REMARKS

I. Explanation of Amendments

The new claims are largely based on the original claims but rewritten for ease of reading. The alleles of particular polymorphisms recited in the claims are supported in the specification at page 50, lines 16-23 and in Table 7 at page 87. The new claims do not add new matter to the application. The Applicants canceled claims 1-60 without prejudice and do not intend by these amendments to abandon the subject matter of any claim, and reserve the right to pursue claims to any invention that is described in this application in related applications, such as continuing applications.

II. Remarks Relating to the Restriction Requirement

Both the restriction requirement and various rejections are based on a fundamental misunderstanding of the claims. For example, the Patent Office justifies the continued application of the restriction requirement on the notion that "there is no common utility" and that "the claims are drawn to the differences, *i.e.* the polymorphisms in the FLAP gene, and not the common structural features of the FLAP gene." The elected claims are not drawn to polynucleotides, to polymorphisms, or to "differences." Rather, the claims are drawn to methods that involve analyzing a human individual's DNA at a particular locus. The results of the analysis determine whether or not the individual is scored as having elevated risk for myocardial infarction. There is common utility for all variations of this method that are described in the application.

III. The Rejection Under 35 U.S.C. § 112, First paragraph for Lack of Adequate Written Description Should be Withdrawn

Claims 1 and 2 were rejected under 35 U.S.C. § 112, first paragraph for allegedly lacking an adequate written description. The Examiner asserted that the claims encompass a large genus of nucleic acids which comprise variants in any region and the specification fails to describe the common attributes or characteristics that identify members of the claimed genus.

The written description requirement focuses on the invention that is actually claimed in the patent application. See, e.g., *Vas-Cath*, cited by the Patent Office at page 4 of the Office action. The elected claims are directed to a method of assessing susceptibility to

myocardial infarction or stroke in a human individual. As such, the written description analysis should evaluate whether *the claimed method* is adequately described. To the extent that the Patent Office has focused its attention on whether the application contains "an adequate written description of a DNA," the Patent Office is committing legal error. There is no obligation to describe a DNA invention when the present claims are directed to a method.

Notwithstanding this error, the Applicant acknowledges with thanks the Patent Office's determination that the previous presented claims that refer to specific polymorphisms are adequately described. The new claims 61-66 continue to refer to a FLAP haplotype comprising specific polymorphisms, and the basis for rejection continues to be inapplicable to these claims.

In view of the foregoing remarks, the newly presented claims are adequately described. Therefore, the rejection under 35 U.S.C. § 112, first paragraph for lack of written description should be withdrawn.

IV. The Rejection Under 35 U.S.C. § 112, First Paragraph for Lack of Enablement Should be Withdrawn

Claims 1-4 and 33-60 were rejected under 35 U.S.C § 112, first paragraph for allegedly lacking enablement. The Examiner stated that due to the unpredictability in the art and the broad genus claims, one of skill in the art would be required to perform an undue amount of experimentation to make and use the claimed invention. Applicants traverse this rejection.

To determine if the claims require undue experimentation, the factors set out in *In re Wands et al.*, 858 F2d. 731, 737, 8 USPQ 2d 1400, 1404 (Fed. Circ. 1988) are considered including: 1) the breadth of the claims, 2) the nature of the invention, 3) state of the prior art, 4) the level of one of ordinary skill, 5) the level of predictability in the art, 6) the amount of direction provided by the inventor, 7) the existence of working examples and 8) quantity of examination needed to make or use the invention based on the content of the disclosure.

A. Nature of the Invention and Breath of the Claims

The nature of the invention is methods of assessing susceptibility to myocardial infarction (MI) or stroke comprising screening nucleic acid of a human individual

to determine whether the nucleic acid has a particular haplotype comprised of certain enumerated polymorphisms in a FLAP nucleic acid. Therefore, the nature of the invention and the breath of the amended claims does not encompass any polymorphism in any FLAP nucleic acid as asserted by the Examiner. The application suitably describes DNA/RNA manipulation and analytical techniques for practicing the invention.

The data provided in the specification and the declaration are the results of an association study which determined that there was a statistically significant relative risk associated with the presence of haplotypes in the FLAP nucleic acid sequence. The haplotypes of the invention are associated with relative risk for developing MI or stroke and have not been determined or asserted to be the cause of that risk. The data provided in the specification, in particular Tables 5, 6, 7, 8 and 9 at pages 84-90, and the data presented in the Declaration of Anna Helgadottir, M.D. under 37 C.F.R. § 1.132 submitted herewith (denoted herein as "the Declaration," attached as Exhibit 1) demonstrate that detection of the claimed FLAP haplotypes are robust predictors of an increased risk of developing MI or stroke.

The method claims of the invention relate to assessing the susceptibility of a human individual for developing MI or stroke. These methods are diagnostic screens that will identify a target population. In modern medicine, the ultimate goal is to design the perfect diagnostic test, but this lofty goal is rarely, if ever, achieved with known diagnostic tests. For example, there is not a perfect correlation between cholesterol level and cardiac disease but the medical community continues to regularly test cholesterol levels and prescribe treatments to lower cholesterol levels because it is considered a risk-factor for cardiac disease. In addition, screening for alterations in the breast cancer genes, BRAC1 and BRCA2 is regularly, performed in women who may be at risk of developing breast cancer, even though only 36 to 85 percent (360–850 out of 1,000) of women with an altered BRCA1 or BRCA2 gene will develop breast cancer. In addition, current clinical trials include using diagnostic assays as an indicator that a therapy is likely to be effective in a particular patient population. For example, the Trofile diagnostic assay is a co-receptor trophism assay that identifies whether an individual strain of HIV uses CCR5 co-receptor, CXCR4 co-receptor or both co-receptors to infect healthy cells. This assay currently is being tested as a screening method for determining if a patient is likely to respond to treatment with a CCR5 antagonist. This screening is useful but not perfect because even though a HIV virus uses CCR5 to infect

cells in 80% of early infections, the virus can mutate and use CCR4 or both co-receptors as the infection progresses. (See Exhibit 2).

The claimed diagnostic methods may not be perfect but these methods are enabled because the specification teaches a correlation of the disclosed polymorphisms and haplotypes with a significant relative risk for developing MI or stroke. Therefore, the specification teaches how to make and use the claimed methods and these claimed methods are medically useful even if not definitive.

B. Unpredictability in the Art

The Examiner stated that the art teaches that ethnicity-specific risk of myocardial infarction (MI) in the ALOX5AP and FLAP gene is unpredictable. The Examiner cited Helgadottir *et al.*, (*Nat. Genetics* 38: 68-74, 2006), Meschia *et al.* (*Ann. Neurology* 58: 351-361, 2005), Hirschhorn *et al.* (*Genet. Med.* 4: 45-61, 2002), Ioannidis *et al.* (*Nat. Genetics* 29: 306-309, 2001) and Meyer *et al.* U.S. Patent Publication No. 2003/0092019) as evidence of the asserted unpredictability. Submitted herewith is a Declaration of Anna Helgadottir M.D. under 37 C.F.R. § 1.132 (Exhibit 1), which provides further evidence that the studies described in the specification can be (and have been) replicated in different populations. In addition, the declaration refutes the evidence provided by the Examiner to support the asserted unpredictability in the art.

Table 1 in the Declaration provides the association analysis of FLAP haplotype HapA (SG13S99, allele T; SG13S25, allele G; SG13S114, allele T; SG13S89, allele G; SG13S32, allele A) for 7 different population in which a total of 18107 individuals were analyzed. The relative risk (rr) was greater than 1 for six out of the seven populations with an overall risk of 1.12. The data had a collective *P* value of 0.003. In addition, the table summarizes an association analysis (published by a different research group) that substantiates the work of the invention. This data is strong evidence that the association of the claimed haplotypes to increased relative risk for developing MI or stroke is reproducible and predictable.

The Examiner stated that Helgadottir *et al.*, (2006) demonstrates that a haplotype (HapK) had varying degrees of relative risk for MI in different ethnicities. However, Helgadottir *et al.* (2006) discloses association data for the Leukotriene A4

Hydrolase (LTA4H), a member of the leukotriene pathway, but a gene distinct from the FLAP gene. The HapK haplotype comprises markers over the LTA4H gene, not FLAP. This reference is therefore not relevant for assessing the predictability of the present invention, although it tend to reinforce the position that the leukotriene pathway (in which both genes act) has relevance to MI.

The Examiner also stated that Meschia *et al.* demonstrates that the haplotype HapA was not associated with risk for stroke in a British population and there is no evidence supporting linkage of ALOX5AP or PDE4D with stroke. As described in the Declaration, the study in Meschia *et al.* used a small sample of 104 sibling pairs, used a linkage analysis rather than an association analysis, and only showed association results for single markers rather than haplotype association results (See paragraphs 5 and 6 of the Declaration). The current claims require the detection of a FLAP haplotype and the screening methods are based on association studies that were carried out in a large population of patients (over 18000, see Table 1 of the Declaration.). Furthermore, Meschia *et al.* admits that their study might not have possessed sufficient power to detect minor effects of a haplotype (see page 358, left column). Therefore, the study described in Meschi *et al.* does not undermine the results provided in the specification and the Declaration.

Two articles (Hirschhorn *et al.* and Ioannidis *et al*) that revealed the results of genetic association studies were cited by the Examiner to demonstrate that genetic variations are often not reproducible and the variations may be overestimated to correlate with a disease. Hirschhorn *et al.* reviewed 166 genetic associations to determine whether subsequent studies on the same polymorphism and disease also reached statistical significance. In their analysis, only 6 of the associations have been consistently replicated. It should be noted that 97/166 of association were, if fact, observed again in one or more studies. The Examiner stated Hirschhorn *et al* cautions in drawing conclusion from a single report. As demonstrated in the specification and the Declaration, the association of FLAP haplotypes with risk for MI or stroke has been carried out in a more than one study using different populations. In addition, Hirschhorn *et al.* also suggests solutions that could remedy the observed irreproducibility. This article is evidence that those of skill in the art at the time of filing understood what is needed to properly carry out genetic association studies. The Applicant's data is superior to the questionable studies in Hirchhorn *et al.*

Ioannidis *et al.* compared the analysis of 370 genetic studies. The results of this analysis cautioned that a strong association in the first study typically becomes gradually less prominent as more data accumulates. However, the analysis in Ioannidis *et al.* also revealed that in some studies, a first analysis did not find a statistically significant difference but with the accumulation of further data, the genetic association become formally statistically significant (see page 307, left column). The association FLAP haplotypes with risk for MI and stroke were repeated and are not the result of a single small study. As shown in Table 1 of the Declaration, as the sample size increased, collectively the association become more significant (the cumulative *P* value was 0.003), while the relative risk of the different cohorts and the cumulative population remained similar. Therefore, while Ioannidis *et al.* provides evidence that genetic studies may be unpredictable, if the assays are repeated with large populations, a truly significant result may be obtained. As stated in the Declaration "if an association is observed in several studies, it is likely to represent a significant finding." (See paragraph 4 of the Declaration).

The Examiner also cited to Meyer *et al.* to demonstrate that the association of a single SNP in a gene does not indicate that all SNPs with the gene are associated with the disease. The amended claims specify screening for the presence of a haplotype comprising multiple polymorphisms to indicate a risk for MI or stroke. The generalities for which Meyer *et al.* is cited are not relevant to the specific haplotypes recited in the current claims.

C. Guidance in the Specification and Quantity of Experimentation

The Examiner stated that the specification provides no guidance that a skilled artisan could practice the claimed invention as broadly claimed. The amended claim set detects a FLAP haplotype that comprises at least polymorphisms SG13S114, allele T; SG13S32, allele A; SG13S25, allele G; and SG13S89, allele G. The data set out in Tables 5, 6, 7, 8 and 9 (pages 84-90) demonstrate that these haplotypes are associated with significant relative risk for developing MI or stroke. Additional data is provided in the Declaration which demonstrates that these association are reproducible. In addition, the specification teaches one of skill the methods for determining whether a known haplotype is associated with risk for a disease state. Therefore, the specification provides adequate guidance for carrying out the claimed invention.

Even in unpredictable arts, a disclosure of every operable species is not required (See MPEP § 2164.03). In the chemical arts, patents are regularly granted on a genus of compounds even though the specification only provides experimental data on a few species. In addition, the amount of experimental data required for FDA approval is not a prerequisite for patenting a new drug or chemical composition. It is not the role of the U.S. Patent Office to determine that a drug, or in this instance a diagnostic test, is commercially useful and safe. (See *In re Anthony* 56 C.C.P.A. 1443, 1457, 1969). As discussed in detail above, the specification discloses the claimed FLAP haplotypes that are associated with significant relative risk for developing MI or stroke and provides working examples to demonstrate this association. In addition, the Declaration provides evidence that this association is observed in a number of different populations.

The Examiner indicated that the term "individual" includes any animal in addition to humans. The amended claims are directed to recite "human individual" Therefore, the Examiner's concerns regarding whether the polymorphisms are conserved among mammals is now moot.

D. Conclusion

In view of the foregoing remarks and the evidence submitted in the Declaration, claims 61-66 are enabled. Applicants request that the rejection under 35 U.S.C. § 112, first paragraph for lack of enablement be withdrawn.

V. The Rejection Under 35 U.S.C. § 112, Second Paragraph Should be Withdrawn

Claim 2 was rejected under 35 U.S.C. § 112, second paragraph as being indefinite. In the foregoing amendment claim 2 was canceled without prejudice, and therefore this rejection is moot.

CONCLUSION

In view of the foregoing amendment and remarks, Applicants believe pending claims 61-66 are in condition for allowance and early notice thereof is solicited.

Dated: January 22, 2007

Respectfully submitted,

By 

David A. Gass
Registration No.: 38,153
MARSHALL, GERSTEIN & BORUN LLP
233 S. Wacker Drive, Suite 6300
Sears Tower
Chicago, Illinois 60606-6357
(312) 474-6300
Attorney for Applicant

nature genetics

The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke

Anna Helgadottir[1], Andrei Manolescu[1], Gudmar Thorleifsson[1], Solveig Gretarsdottir[1], Helga Jonsdottir[1], Unnur Thorsteinsdottir[1], Nilesh J Samani[2], Gudmundur Gudmundsson[1], Struan F A Grant[1], Gudmundur Thorgeirsson[3], Sigurlaug Sveinbjornsdottir[3], Einar M Valdimarsson[3], Stefan E Matthiasson[3], Halldor Johannsson[3], Olof Gudmundsdottir[1], Mark E Gurney[1], Jesus Sainz[1], Margret Thorhallsdottir[1], Margret Andresdottir[1], Michael L Frigge[1], Eric J Topol[4], Augustine Kong[1], Vilmundur Gudnason[5], Hakon Hakonarson[1], Jeffrey R Gulcher[1] & Kari Stefansson[1]

We mapped a gene predisposing to myocardial infarction to a locus on chromosome 13q12–13. A four-marker single-nucleotide polymorphism (SNP) haplotype in this locus spanning the gene *ALOX5AP* encoding 5-lipoxygenase activating protein (FLAP) is associated with a two times greater risk of myocardial infarction in Iceland. This haplotype also confers almost two times greater risk of stroke. Another *ALOX5AP* haplotype is associated with myocardial infarction in individuals from the UK. Stimulated neutrophils from individuals with myocardial infarction produce more leukotriene B4, a key product in the 5-lipoxygenase pathway, than do neutrophils from controls, and this difference is largely attributed to cells from males who carry the at-risk haplotype. We conclude that variants of ALOX5AP are involved in the pathogenesis of both myocardial infarction and stroke by increasing leukotriene production and inflammation in the arterial wall.

Cardiovascular diseases (CVD) are the leading causes of death and disability in the developed world[1], with an increasing prevalence due to the aging of the population and the obesity epidemic. More than 1 million deaths in the US alone were caused by myocardial infarction and stroke in 2003 (ref. 2). Some of the processes underlying myocardial infarction are now understood: it is generally attributed to atherosclerosis with arterial wall inflammation that ultimately leads to plaque rupture, fissure or erosion[3,4]. This process is known to involve diapedesis of monocytes across the endothelial barrier; activation of neutrophils, macrophage cells and platelets; and release of a variety of cytokines and chemokines[5,6], but the genetic basis of the process has not yet been deciphered.

Two different approaches have been used to search for genes associated with myocardial infarction. SNPs in candidate genes have been tested for association and have, in general, not been replicated or confer only a modest risk of myocardial infarction. Case-control association studies have identified several proinflammatory genes with variants that are associated with either an increased risk of myocardial infarction or a protective effect[7–9]. Four genome-wide scans in families with myocardial infarction have yielded several loci with formidable linkage peaks, but the gene(s) underlying these loci have not yet been identified[10–14]. In addition, one large pedigree study identified a deletion mutation of a transcription factor gene, *MEF2A*, with autosomal dominant transmission[14]. This is an interesting cause of myocardial infarction, but the prevalence of this or other mutations in *MEF2A* outside this family remains to be determined.

Here we report a genome-wide scan of 296 multiplex Icelandic families including 713 individuals with myocardial infarction. Through suggestive linkage to a locus on chromosome 13q12–13, we identified the gene (*ALOX5AP*) encoding FLAP and found that a four-SNP haplotype in the gene confers a nearly two times greater risk of myocardial infarction and stroke. FLAP is a regulator[15] of a crucial pathway in the genesis of leukotriene inflammatory mediators, which are implicated in atherosclerosis both in a mouse model[16] and in human studies[17,18]. Males had the strongest association to the at-risk haplotype, and male carriers of the at-risk haplotype also had significantly greater production of leukotriene-B4 (LTB4), supporting the idea that proinflammatory activity has a role in the pathogenesis of myocardial infarction. We confirmed the association of *ALOX5AP* with myocardial infarction in an independent cohort of British individuals with another haplotype. These results indicate that *ALOX5AP* is the first specific gene isolated that confers substantial population-attributable risk (PAR) of the complex traits of both myocardial infarction and stroke.

[1]deCODE genetics, Sturlugata 8, Reykjavik, Iceland. [2]Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, UK. [3]National University Hospital, Reykjavik, Iceland. [4]Cleveland Clinic Foundation, Cleveland, Ohio, USA. [5]Icelandic Heart Association, Reykjavik, Iceland. Correspondence should be addressed to K.S. (kstefans@decode.is).

Published online 8 February 2004; doi:10.1038/ng1311

EXHIBIT A

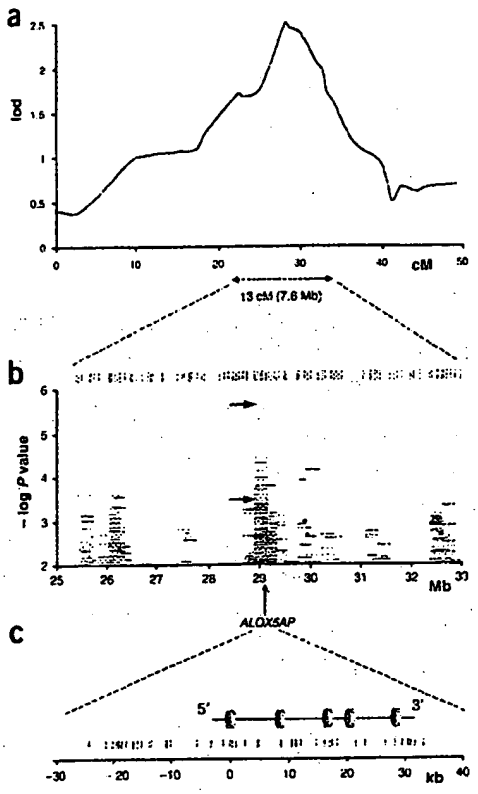


Figure 1 Schematic view of the chromosome 13 linkage region showing *ALOX5AP*. (**a**) The linkage scan for females with myocardial infarction and the one-lod drop region that includes *ALOX5AP*. (**b**) Microsatellite association for all individuals with myocardial infarction: single-marker association (black dots) and two-, three-, four- and five-marker haplotype association (black, blue, green and red horizontal lines, respectively). The blue and red arrows indicate the location of the most significant haplotype association across *ALOX5AP* in males and females, respectively. (**c**) *ALOX5AP* gene structure, with exons shown as colored cylinders, and the locations of all SNPs typed in the region. The green vertical lines indicate the position of the microsatellites (**b**) and SNPs (**c**) used in the analysis.

RESULTS

Linkage analysis

We carried out a genome-wide scan in search of myocardial infarction susceptibility genes using a framework set of 1,068 microsatellite markers. The initial linkage analysis included 713 individuals with myocardial infarction who fulfilled the World Health Organization (WHO) MONICA research criteria[19] and were clustered in 296 extended families. We repeated the linkage analysis for individuals with early onset, for males and for females separately. A description of the number of affected individuals and families in each analysis is provided in **Supplementary Table 1** online, and the corresponding allele-sharing lod scores are given in **Supplementary Figure 1** online. None of these analyses yielded a locus of genome-wide significance. The most promising lod score (2.86) was observed on chromosome 13q12–13 for linkage with females with myocardial infarction at the peak marker *D13S289* (**Supplementary Fig. 1** online). This locus also had the most promising lod score (2.03) for individuals with early-onset myocardial infarction. After we increased the information on identity-by-descent sharing to over 90% by typing an additional 14 microsatellite markers in a 30-cM region around *D13S289*, the lod score for the association in females dropped to 2.48 ($P = 0.00036$), and the lod score remained highest at *D13S289* (**Fig. 1a**). In an independent linkage study of males with ischemic stroke or transient ischemic attack (TIA), we observed linkage to the same locus with a lod score of 1.51 at the same peak marker (**Supplementary Fig. 2** online), further suggesting that a cardiovascular susceptibility factor might reside at this locus.

Microsatellite association study

The 7.6-Mb region that corresponds to a drop of 1 in lod score in the female–myocardial infarction linkage analysis contains 40 known genes (**Supplementary Table 2** online). To determine which gene in this region was most likely to contribute to myocardial infarction, we typed 120 microsatellite markers in the region and carried out a case-control association study using 802 unrelated (separated by at least three meioses) individuals with myocardial infarction and 837 population-based controls. We also repeated the association study for each of the three phenotypes that were used in the linkage study: individuals with early onset, males and females with myocardial infarction. In addition to testing each marker individually, we also tested haplotypes based on these markers for association. To limit the number of haplotypes tested, we considered only haplotypes spanning less than 300 kb that were over-represented among the affected individuals.

The haplotype with the strongest association to myocardial infarction ($P = 0.00004$) covered a region that contains two known genes: *ALOX5AP* (**Fig. 1b**) and a gene with an unknown function called highly charged protein (*D13S106E*). The haplotype association in this region for females with myocardial infarction was less significant ($P = 0.0004$) than for all individuals with myocardial infarction, and the most significant haplotype association was observed for males with myocardial infarction ($P = 0.000002$). The haplotype associated with males with myocardial infarction was the only haplotype that retained significant association after adjusting for all haplotypes tested.

FLAP, together with 5-lipoxygenase (5-LO), is a regulator of the leukotriene biosynthetic pathway that has recently been implicated in the pathogenesis of atherosclerosis[16–18]. Therefore, *ALOX5AP* was a good candidate for the gene underlying the association with myocardial infarction.

Screening for SNPs in *ALOX5AP* and LD mapping

To determine whether variations in *ALOX5AP* significantly associate with myocardial infarction and to search for causal variations, we sequenced *ALOX5AP* in 93 affected individuals and 93 controls. The sequenced region covers 60 kb containing *ALOX5AP*, including the five known exons and introns, the 26-kb region 5′ to the first exon and the 7-kb region 3′ to the fifth exon. We identified 144 SNPs, of which we excluded 96 from further analysis owing to either a low minor allele frequency or complete correlation (redundancy) with other SNPs. **Figure 1c** shows the distribution of the 48 SNPs chosen for genotyping, relative to exons, introns and the 5′ and 3′ flanking regions of *ALOX5AP*. We identified only one SNP in a coding sequence (exon 2), which did not lead to an amino acid substitution. The locations of the 48 SNPs in the National Center for Biotechnology Information human genome assembly build 34 are listed in **Supplementary Table 3** online. In addition to the SNPs, we typed a polymorphism consisting of a monopolymer A repeat in the *ALOX5AP* promoter region[20].

The linkage disequilibrium (LD) block structure defined by the 48 genotyped SNPs is shown in **Figure 2**. Strong LD was detected across the *ALOX5AP* region, although at least one historical recombination seems to have occurred, dividing the region into two strongly correlated LD blocks.

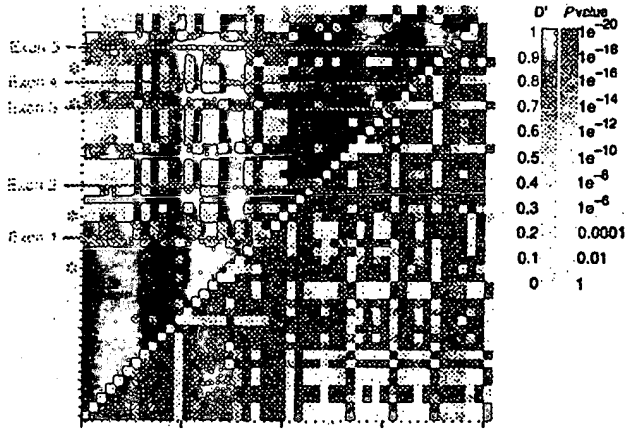


Figure 2 Pairwise LD between SNPs in a 60-kb region encompassing *ALOX5AP*. The markers are plotted equidistantly. Two measures of LD are shown: *D'* in the upper left triangle and *P* values in the lower right triangle. Colored lines indicate the positions of the exons of *ALOX5AP*, and the green stars indicate the location of the markers of the at-risk haplotype HapA. Scales for both measures of the LD strength are provided on the right.

Haplotype association with myocardial infarction

In a case-control association study, we genotyped the 48 selected SNPs and the monopolymer A repeat marker in a set of 779 unrelated individuals with myocardial infarction and 624 population-based controls. We tested each of the 49 markers individually for association with the disease. Three SNPs, one located 3 kb upstream of the first exon and the other two 1 kb and 3 kb downstream of the first exon, showed nominally significant association to myocardial infarction (**Supplementary Table 4** online). After adjusting for the number of markers tested, however, these results were not significant. We then searched for haplotypes associated with the disease using the same cohorts. We limited the search to haplotype combinations constructed from two, three or four SNPs and tested only haplotypes that were over-represented in the individuals with myocardial infarction. The resulting *P* values were adjusted for all the haplotypes we tested by randomizing the affected individuals and controls.

Several haplotypes were significantly associated with the disease at an adjusted significance level of $P < 0.05$ (**Supplementary Table 5** online). We observed the most significant association with a four-SNP haplotype spanning 33 kb, including the first four exons of *ALOX5AP* (**Fig. 1c**), with a nominal *P* value of 0.0000023 and an adjusted *P* value of 0.005. This haplotype, called HapA, has a haplotype frequency of 15.8% (carrier frequency 29.1%) in affected individuals versus 9.5% (carrier frequency 18.1%) in controls (**Table 1**). The relative risk conferred by HapA compared with other haplotypes constructed from the same SNPs, assuming a multiplicative model, was 1.8 and the corresponding PAR was 13.5%. HapA was present at a higher frequency in males (carrier frequency 30.9%) than in females with myocardial infarction (carrier frequency 25.7%; **Table 1**). All other haplotypes that were significantly associated with an adjusted *P* value less than 0.05 were highly correlated with HapA and should be considered variants of that haplotype (**Supplementary Table 5** online).

Association of HapA with stroke and PAOD

Because of the high degree of comorbidity among myocardial infarction, stroke and peripheral arterial occlusive disease (PAOD), with most of these cases occurring on the basis of an atherosclerotic disease, we wanted to determine whether HapA was also associated with stroke or PAOD. We typed the SNPs defining HapA for these cohorts. We removed first- and second-degree relatives and all known cases of myocardial infarction and tested for association in 702 individuals with stroke and 577 individuals with PAOD (**Table 1**). We observed a significant association of HapA with stroke, with a relative risk of 1.67 ($P = 0.000095$). In addition, we determined whether HapA was primarily associated with a particular subphenotype of stroke and found that both ischemic and hemorrhagic stroke were significantly associated with HapA (**Supplementary Table 6** online). Finally, although HapA was more frequent in the PAOD cohort than in the population controls (**Table 1**), this was not significant. Similar to the stronger association of HapA with males with myocardial infarction than with females with myocardial infarction, HapA also showed stronger association with males than with females with stroke and PAOD (**Table 1**).

Haplotype association in a British cohort

In an independent study, we determined whether variants in *ALOX5AP* also affected the risk of myocardial infarction in a population outside Iceland. We typed SNPs defining HapA in a cohort of 753 individuals from the UK who had sporadic myocardial infarction and in 730 British population controls. The affected individuals and controls were from three separate study cohorts recruited in Leicester and Sheffield. We found a slightly higher frequency of HapA in affected individuals versus controls (16.8% versus 15.1%, respectively), but the results were not statistically significant. As in the Icelandic population, HapA was more common in males with myocardial infarction (carrier frequency 31.7%) than in females with myocardial infarction (carrier frequency 28.0%). When we typed an additional nine SNPs, distributed across *ALOX5AP*, in the British cohort and searched for other haplotypes that might be associated with myocardial infarction, two SNPs showed association to myocardial infarction with a nominally significant *P* value (data not shown). Moreover, three- and four-SNP haplotype combinations were associated with higher risk of myocardial infarction in the British cohort, and we observed the most signifi-

Table 1 Association of HapA with myocardial infarction, stroke and PAOD

Phenotype (*n*)	Frequency	RR	PAR	*P* value	*P* value[a]
Myocardial infarction (779)	0.158	1.80	0.135	0.0000023	0.005
Males (486)	0.169	1.95	0.158	0.00000091	ND
Females (293)	0.138	1.53	0.094	0.0098	ND
Early onset (358)	0.139	1.53	0.094	0.0058	ND
Stroke (702)[b]	0.149	1.67	0.116	0.000095	ND
Males (373)	0.156	1.76	0.131	0.00018	ND
Females (329)	0.141	1.55	0.098	0.0074	ND
PAOD (577)[b]	0.122	1.31	0.056	0.061	ND
Males (356)	0.126	1.36	0.065	0.057	ND
Females (221)	0.114	1.22	0.041	0.31	ND

[a]*P* value adjusted for the number of haplotypes tested. [b]Excluding known cases of myocardial infarction.

Shown is HapA of *ALOX5AP* and the corresponding number of affected individuals (*n*), the haplotype frequency in affected individuals, the relative risk (RR), PAR and *P* values. HapA is defined by the SNPs SG13S25, SG13S114, SG13S89 and SG13S32 (**Supplementary Table 5** online). The same controls ($n = 624$) were used for the association analysis in myocardial infarction, stroke and PAOD as well as for the analysis of males, females and individuals with early onset. The frequency of HapA in the control cohort is 0.095. ND, not done.

Table 2 Association of HapB with myocardial infarction in British individuals

Phenotype (*n*)	Frequency	RR	PAR	*P* value	*P* value[a]
Myocardial infarction (753)	0.075	1.95	0.072	0.00037	0.046
Males (549)	0.075	1.97	0.072	0.00093	ND
Females (204)	0.073	1.90	0.068	0.021	ND

[a]*P* value adjusted for the number of haplotypes tested using 1,000 randomization tests.

Shown are the results for HapB that shows the strongest association in the British myocardial infarction cohort. HapB is defined by the SNPs SG13S377, SG13S114, SG13S41 and SG13S35, which have the alleles A, A, A and G, respectively. In all three phenotypes shown, the same set of 730 British controls was used and the frequency of HapB in the control cohort is 0.040. Number of affected individuals (*n*), haplotype frequency in affected individuals, relative risk (RR) and PAR are indicated. ND, not done.

cant association for a four-SNP haplotype with a nominal *P* value of 0.00037 (Table 2). We call this haplotype HapB. The haplotype frequency of HapB was 7.5% in the individuals with myocardial infarction (carrier frequency 14.4%) compared with 4.0% (carrier frequency 7.8%) in controls, conferring a relative risk of 1.95 (Table 2). This association of HapB remained significant after adjusting for all haplotypes tested, using 1,000 randomization steps, with an adjusted $P = 0.046$. No other SNP haplotype had an adjusted *P* value <0.05. The two at-risk haplotypes, HapA and HapB, are mutually exclusive; there are no instances in which the same chromosome carries both haplotypes.

More LTB4 in individuals with myocardial infarction

To determine whether individuals with a past history of myocardial infarction had greater activity of the 5-LO pathway than controls, we measured production of LTB4 (a key product of the 5-LO pathway) in blood neutrophils isolated from Icelandic individuals with myocardial infarction and controls before and after stimulation with the calcium ionophore ionomycin. We detected no difference in LTB4 production in resting neutrophils from individuals with myocardial infarction versus controls. In contrast, LTB4 generation by neutrophils stimulated with ionomycin was substantially greater in individuals with myocardial infarction than in controls after 15 and 30 min, respectively (**Fig. 3a**). Moreover, the observed difference in release of LTB4 was largely accounted for by male carriers of HapA (**Fig. 3b**), whose cells produced significantly more LTB4 than cells from controls ($P = 0.0042$; **Supplementary Table 7** online). There was also a heightened LTB4 response in males who did not carry HapA, but this difference was of borderline significance (**Supplementary Table 7** online). This could be explained by additional variants in *ALOX5AP* that have not been uncovered, or in other genes belonging to the 5-LO pathway, that may account for upregulation of the LTB4 response in some individuals without the *ALOX5AP* at-risk haplotype. We did not detect differences in LTB4 response in females (**Supplementary Table 7** online), but because of the small sample size, this result is not conclusive. The elevated levels of LTB4 production in stimulated neutrophils from male carriers of the at-risk haplotype suggest that the disease-associated variants of *ALOX5AP* heighten the response of FLAP to factors that stimulate inflammatory cells.

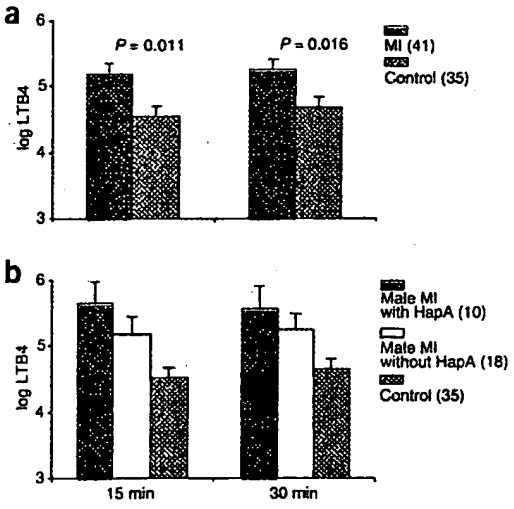


Figure 3 LTB4 production of ionomycin-stimulated neutrophils from individuals with myocardial infarction ($n = 41$) and controls ($n = 35$). The log-transformed (mean ± s.d.) values measured at 15 and 30 min in stimulated cells are shown. (**a**) LTB4 production in individuals with myocardial infarction (MI) and controls. The difference in the mean values between affected individuals and controls was tested using a two-sample *t*-test of the log-transformed values. (**b**) LTB4 production in males with myocardial infarction carrying HapA (red bars) and not carrying HapA (white bars). Mean values of controls (blue bars) are included for comparison. Males with HapA produced the highest amounts of LTB4 ($P < 0.005$ compared with controls). Data for females are shown in **Supplementary Table 7** online.

DISCUSSION

Our results show that variants of *ALOX5AP* encoding FLAP are associated with greater risk of myocardial infarction and stroke. In our Icelandic cohort, a haplotype that spans *ALOX5AP* is carried by 29.1% of all individuals with myocardial infarction and almost doubles the risk of myocardial infarction. We then replicated these findings in an independent cohort of individuals with stroke. Furthermore, stimulated neutrophils from individuals with myocardial infarction had greater production of LTB4, one of the key products of the 5-LO pathway. When we examined this in the context of the at-risk haplotype, however, the gain of function was largely attributed to male carriers of the at-risk haplotype, who also had the strongest association with the *ALOX5AP* haplotype. Another haplotype spanning *ALOX5AP* was associated with myocardial infarction in a British cohort. Although the pathogenic variants responsible for the effects associated with the disease haplotypes are unknown, the greater production of LTB4 observed in ionomycin-stimulated neutrophils from male carriers of the at-risk haplotype suggests that the disease-associated variants increase the response of FLAP to factors that stimulate inflammatory cells.

We observed suggestive linkage to chromosome 13q12–13 with several different phenotypic groups, including females with myocardial infarction, individuals of both sexes with early-onset myocardial infarction and males with ischemic stroke or TIA. But we observed the strongest haplotype association for males with myocardial infarction or stroke. Therefore, the linkage signal in females with myocardial infarction and in individuals with early-onset myocardial infarction is not explained by the at-risk haplotype that we identified, and we expect that there may be other unidentified variants or haplotypes in *ALOX5AP*, or in other genes in the linkage region, that may confer risk of these cardiovascular phenotypes. These variants are probably rarer than HapA with relatively high penetrance, higher in women than in men.

FLAP has an important role in the initial steps of leukotriene biosynthesis[15], which is largely confined to leukocytes and can be

triggered by a variety of stimuli. In this biosynthetic pathway, unesterified arachidonic acid is converted to LTA4 by the action of 5-LO and its activating protein FLAP. The unstable epoxide LTA4 is further metabolized to LTB4 or LTC4 by LTA4 hydrolase and LTC4 synthase, respectively. In addition, LTA4 can be exported to neighboring cells that are devoid of 5-LO activity and become subject to transcellular leukotriene biosynthesis[21–23]. The leukotrienes have a variety of proinflammatory effects[24,25]. LTB4 activates leukocytes, leading to chemotaxis and increased adhesion of leukocytes to vascular endothelium, release of lysosomal enzymes such as myeloperoxidase and production of superoxide anions[25]. The cysteinyl-containing leukotrienes (LTC4 and its metabolites LTD4 and LTE4) increase vascular permeability in postcapillary venules and are potent vasoconstrictors of coronary arteries[26–28].

The importance of the 5-LO pathway is well established in asthma, and drugs inhibiting this pathway have been developed for treating asthma. The role of the 5-LO pathway in the pathogenesis of atherosclerosis has recently received attention. A study of postmortem pathologic specimens showed an increase in the expression of members of the 5-LO pathway, including 5-LO and FLAP, in atherosclerotic lesions at various stages of development in the aorta, coronary arteries and carotid arteries[18]. Furthermore, 5-LO was localized to macrophages, dendritic cells, foam cells, mast cells and neutrophilic granulocytes, and the number of cells expressing 5-LO was markedly greater in advanced lesions[18]. The leukocytes positive for 5-LO accumulated at distinct sites that are most prone to rupture[29], such as the shoulder regions below the fibrous cap of the atherosclerotic lesion[18]. A 5-LO promoter variant is associated with abnormal carotid artery intima-media thickness and heightened inflammatory biomarkers[30]. In addition, antagonists of LTB4 block the development of atherosclerosis in apo-E-deficient and LDRL-deficient mice[31], and a congenic mouse strain with a heterozygous deficiency of 5-LO shows resistance to atherosclerosis[16], further supporting the idea that greater activity of the 5-LO pathway has a role in predisposition to atherosclerosis.

Our data also show that the at-risk haplotype of *ALOX5AP* has higher frequency in all subgroups of stroke, including ischemic stroke, TIA and hemorrhagic stroke. HapA confers significantly higher risk of myocardial infarction and stroke than it does of PAOD. This could be explained by differences in the pathogenesis of these diseases. Unlike individuals with PAOD, who have ischemic legs because of atherosclerotic lesions that are responsible for gradually diminishing blood flow to the legs, individuals with myocardial infarction and stroke have suffered acute events, with disruption of the vessel wall suddenly decreasing blood flow to regions of the heart and the brain.

We did not find association between HapA and myocardial infarction in a British cohort, but we did find significant association between myocardial infarction and a different *ALOX5AP* variant. The existence of different haplotypes of the gene conferring risk to myocardial infarction in different populations is not unexpected. It is not unreasonable to assume that a common disease like myocardial infarction is associated with many different mutations or sequence variations and that the frequencies of these disease-associated variants may differ between populations. It would also not be unexpected for the same mutation to arise on different haplotypic backgrounds.

Our work suggests that *ALOX5AP* has an important role in the pathogenesis of myocardial infarction and stroke in humans. Our study, together with others, may provide the necessary background to launch therapeutic trials to determine whether pharmacological inhibition of FLAP will prevent the development of myocardial infarction and stroke.

METHODS

Study population. We recruited the individuals in the study from a registry of over 8,000 individuals, which includes all individuals who had myocardial infarctions before the age of 75 in Iceland from 1981 to 2000. This registry is a part of the WHO MONICA Project[19]. Diagnoses of all individuals in the registry follow strict diagnostic rules based on signs, symptoms, electrocardiograms, cardiac enzymes and necropsy findings.

We used genotypes from 713 individuals with myocardial infarction and 1,741 of their first-degree relatives in the linkage analysis. For the microsatellite association study of the locus associated with myocardial infarction, we used 802 unrelated (no first- or second-degree relatives) individuals with myocardial infarction (233 females, 624 males and 302 with early onset) and 837 population-based controls. The females studied were post-menopausal. Over 90% of the individuals were taking aspirin or other nonsteroidal anti-inflammatory drugs. For the SNP association study in and around *ALOX5AP*, we genotyped 779 unrelated individuals with myocardial infarction (293 females, 486 males and 358 with early onset). The control group for the SNP association study was population-based and comprised of 624 unrelated males and females 20–90 years of age whose medical history was unknown. The stroke and PAOD cohorts used in this study have previously been described[32–34]. For the stroke linkage analysis, we used genotypes from 342 males with ischemic stroke or TIA that were linked to at least one other male within and including six meioses in 164 families. For the association studies, we analyzed 702 individuals with all forms of stroke (329 females and 373 males) and 577 individuals with PAOD (221 females and 356 males). Individuals with stroke or PAOD who also had myocardial infarction were excluded. Controls used for the stroke and PAOD association studies were the same as used in the myocardial infarction SNP association study.

The study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. We obtained informed consent from all study participants. Personal identifiers associated with medical information and blood samples were encrypted with a third-party encryption system as previously described[35].

Statistical analysis. We carried out a genome-wide scan as previously described[33], using a set of 1,068 microsatellite markers. We used multipoint, affected-only allele-sharing methods[36] to assess the evidence for linkage. All results were obtained using the program Allegro[37] and the deCODE genetic map[38]. We used the S_{pairs} scoring function[39,40] and the exponential allele-sharing model[36] to generate the relevant 1-degree-of-freedom statistics. When combining the family scores to obtain an overall score, we used a weighting scheme that is halfway on a log scale between weighting each affected pair equally and weighting each family equally. In the analysis, all genotyped individuals who were not affected were treated as 'unknown'. Because of concern with small-sample behavior, we usually computed corresponding *P* values in two different ways for comparison and report the less significant one. The first *P* value was computed based on large sample theory, $Z_{lr} = \sqrt{(2 \log_e(10) \text{ lod})}$, and is distributed approximately as a standard normal distribution under the null hypothesis of no linkage[36]. A second *P* value was computed by comparing the observed lod score with its complete data sampling distribution under the null hypothesis[37]. When a data set consisted of more than a handful of families, these two *P* values tended to be very similar. The information measure we used, which is implemented in Allegro, is closely related to a classical measure of information and has a property that is between 0 (if the marker genotypes are completely uninformative) and 1 (if the genotypes determine the exact amount of allele sharing by descent among the affected relatives)[41,42].

For single-marker association studies, we used Fisher's exact test to calculate two-sided *P* values for each allele. All *P* values are unadjusted for multiple comparisons unless specifically indicated. We present allelic rather than carrier frequencies for microsatellites, SNPs and haplotypes. To minimize any bias due to the relatedness of the individuals who were recruited as families for the linkage analysis, we eliminated first- and second-degree relatives. For the haplotype analysis we used the program NEMO[32], which handles missing genotypes and uncertainty with phase through a likelihood procedure, using the expectation-maximization algorithm as a computational tool to estimate haplotype frequencies. Under the null hypothesis, the affected individuals and controls were assumed to have identical haplotype frequencies. Under the alternative

hypotheses, the candidate at-risk haplotype was allowed to have a higher frequency in the affected individuals than in controls, and the ratios of frequencies of all other haplotypes were assumed to be the same in both groups. Likelihoods were maximized separately under both hypotheses, and a corresponding 1-degree-of-freedom likelihood ratio statistic was used to evaluate statistical significance[32]. Although we only searched for haplotypes that increased the risk, all reported *P* values are two-sided unless otherwise stated. To assess the significance of the haplotype association corrected for multiple testing, we carried out a randomization test using the same genotype data. We randomized the cohorts of affected individuals and controls and repeated the analysis. This procedure was repeated up to 1,000 times, and the *P* value we present is the fraction of replications that produced a *P* value for a haplotype tested that was lower than or equal to the *P* value we observed using the original affected individual and control cohorts.

For both single-marker and haplotype analysis, we calculated relative risk (RR) and PAR assuming a multiplicative model[43,44] in which the risk of the two alleles of haplotypes a person carries multiply. We calculated LD between pairs of SNPs using the standard definition of D' (ref. 45) and R^2 (ref. 46). Using NEMO, we estimated frequencies of the two marker allele combinations by maximum likelihood and evaluated deviation from linkage equilibrium by a likelihood ratio test. When plotting all SNP combinations to elucidate the LD structure in a particular region, we plotted D' in the upper left corner and the *P* value in the lower right corner. In the LD plots we present, the markers are plotted equidistantly rather than according to their physical positions.

Identification of DNA polymorphisms. We identified new polymorphic repeats (dinucleotide or trinucleotide repeats) with the Sputnik program. We subtracted the lower allele of the CEPH sample 1347-02 (CEPH genomics repository) from the alleles of the microsatellites and used it as a reference. We detected SNPs in the gene by PCR sequencing exonic and intronic regions from affected individuals and controls. We also detected public polymorphisms by BLAST search of the National Center for Biotechnology Information SNP database. We genotyped SNPs using a method for detecting SNPs with fluorescent polarization template-directed dye-terminator incorporation[47] and TaqMan assays (Applied Biosystems).

Isolation and activation of peripheral blood neutrophils. We drew 50 ml of blood from each of 41 individuals with myocardial infarction and 35 age- and sex-matched controls into vacutainers containing EDTA. All blood was drawn at the same time in the early morning after 12 h of fasting. We isolated neutrophils using Ficoll-Paque PLUS (Amersham Biosciences).

We collected the red cell pellets from the Ficoll gradient and then lysed red blood cells in 0.165 M ammonium chloride for 10 min on ice. After washing them with phosphate-buffered saline, we counted neutrophils and plated them at 2×10^6 cells ml^{-1} in 4-ml cultures of 15% fetal calf serum (GIBCO BRL) in RPMI-1640 medium (GIBCO BRL). We then stimulated cells with maximum effective concentration of ionomycin (1 μM). At 0, 15, 30, 60 min after adding ionomycin, we aspirated 600 μl of culture medium and stored it at −80 °C for the measurement of LTB4 release as described below. We maintained cells at 37 °C in a humidified atmosphere of 5% carbon dioxide–95% air. We treated all samples with indomethasine (1 μM) to block the cyclooxygenase enzyme.

Ionomycin-induced release of LTB4 in neutrophils. We used the LTB4 Immunoassay (R&D systems) to quantify LTB4 concentration in supernatant from cultured ionomycin-stimulated neutrophils. The assay we used is based on the competitive binding technique in which LTB4 present in the testing samples (200 μl) competes with a fixed amount of alkaline phosphatase–labeled LTB4 for sites on a rabbit polyclonal antibody. During the incubation, the polyclonal antibody becomes bound to a goat antibody to rabbit coated onto the microplates. After washing to remove excess conjugate and unbound sample, a substrate solution was added to the wells to determine the bound enzyme activity. We stopped the color development and read the absorbance at 405 nm. The intensity of the color is inversely proportional to the concentration of LTB4 in the sample. Each LTB4 measurement using the LTB4 Immunoassay was done in duplicate.

British study population. We recruited three separate British cohorts as described previously[48,49]. The first two cohorts comprised 549 individuals from among those who were admitted to the coronary care units of the Leicester Royal Infirmary, Leicester (July 1993–April 1994), and the Royal Hallamshire Hospital, Sheffield (November 1995–March 1997), and satisfied the WHO criteria for acute myocardial infarction in terms of symptoms, elevations in cardiac enzymes or electrocardiographic changes[50]. We recruited 532 control individuals in each hospital from adult visitors of individuals with noncardiovascular disease on general medical, surgical, orthopedic and obstetric wards to find subjects representative of the source population from which the affected individuals originated. Individuals who reported a history of coronary heart disease were excluded.

In the third cohort, we recruited 204 individuals retrospectively from the registries of three coronary care units in Leicester. All had suffered a myocardial infarction according to WHO criteria before the age of 50 years. At the time of participation, individuals were at least 3 months from the acute event. The control cohort comprised 198 individuals with no personal or family history of premature coronary heart disease, matched for age, sex and current smoking status with the cases. We recruited control individuals from three primary care practices located in the same geographical area. In all cohorts, individuals were white of Northern European origin. Local research ethics committees approved all the studies, and individuals provided written informed consent for use of samples in genetic studies of coronary artery disease.

URLs. The Sputnik program is available at http://espressosoftware.com/pages/sputnik.jsp. The National Center for Biotechnology Information SNP database is available at http://www.ncbi.nlm.nih.gov/SNP/index.html.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the affected individuals and their families whose contribution made this study possible and the nurses at the Icelandic Heart Association, personnel at the deCODE core facilities, T. Jonsdottir, F. Runarsson, E. Palsdottir, J. Kostic, K. Channer, R. Steeds, R. Singh and P. Braund for their contributions. N.J.S. is supported by the British Heart Foundation.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Received 7 January; accepted 26 January 2004
Published online at http://www.nature.com/naturegenetics/

1. Bonow, R.O., Smaha, L.A., Smith, S.C. Jr., Mensah, G.A. & Lenfant, C. World Heart Day 2002: the international burden of cardiovascular disease: responding to the emerging global epidemic. *Circulation* **106**, 1602–1605 (2002).
2. *Heart Disease and Stroke Statistics, 2003 Update* (American Heart Association, Dallas, Texas, 2002).
3. Lusis, A.J. Atherosclerosis. *Nature* **407**, 233–241 (2000).
4. Libby, P. Inflammation in atherosclerosis. *Nature* **420**, 868–874 (2002).
5. Stratford, N., Britten, K. & Gallagher, P. Inflammatory infiltrates in human coronary atherosclerosis. *Atherosclerosis* **59**, 271–276 (1986).
6. Poole, J.C. & Florey, H.W. Changes in the endothelium of the aorta and the behaviour of macrophages in experimental atheroma of rabbits. *J. Pathol. Bacteriol.* **75**, 245–251 (1958).
7. Topol, E.J. *et al.* Single nucleotide polymorphisms in multiple novel thrombospondin genes may be associated with familial premature myocardial infarction. *Circulation* **104**, 2641–2644 (2001).
8. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650–654 (2002).
9. Yamada, Y. *et al.* Prediction of the risk of myocardial infarction from polymorphisms in candidate genes. *N. Engl. J. Med.* **347**, 1916–1923 (2002).
10. Broeckel, U. *et al.* A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat. Genet.* **30**, 210–214 (2002).
11. Francke, S. *et al.* A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. *Hum. Mol. Genet.* **10**, 2751–2765 (2001).
12. Harrap, S.B. *et al.* Genome-wide linkage analysis of the acute coronary syndrome suggests a locus on chromosome 2. *Arterioscler. Thromb. Vasc. Biol.* **22**, 874–878 (2002).
13. Pajukanta, P. *et al.* Two loci on chromosomes 2 and X for premature coronary heart disease identified in early- and late-settlement populations of Finland. *Am. J. Hum. Genet.* **67**, 1481–1493 (2000).
14. Wang, L., Fan, C., Topol, S.E., Topol, E.J. & Wang, Q. Mutation of MEF2A in an inherited disorder with features of coronary artery disease. *Science* **302**, 1578–1581 (2003).

15. Dixon, R.A. *et al.* Requirement of a 5-lipoxygenase-activating protein for leukotriene synthesis. *Nature* **343**, 282–284 (1990).
16. Mehrabian, M. *et al.* Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice. *Circ. Res.* **91**, 120–126 (2002).
17. Brezinski, D.A., Nesto, R.W. & Serhan, C.N. Angioplasty triggers intracoronary leukotrienes and lipoxin A4. Impact of aspirin therapy. *Circulation* **86**, 56–63 (1992).
18. Spanbroek, R. *et al.* Expanding expression of the 5-lipoxygenase pathway within the arterial wall during human atherogenesis. *Proc. Natl. Acad. Sci. USA* **100**, 1238–1243 (2003).
19. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. *J. Clin. Epidemiol.* **41**, 105–14 (1988).
20. Koshino, T. *et al.* Novel polymorphism of the 5-lipoxygenase activating protein (FLAP) promoter gene associated with asthma. *Mol. Cell. Biol. Res. Commun.* **2**, 32–35 (1999).
21. Sala, A., Bolla, M., Zarini, S., Muller-Peddinghaus, R. & Folco, G. Release of leukotriene A4 versus leukotriene B4 from human polymorphonuclear leukocytes. *J. Biol. Chem.* **271**, 17944–17948 (1996).
22. Dahinden, C.A., Clancy, R.M., Gross, M., Chiller, J.M. & Hugli, T.E. Leukotriene C4 production by murine mast cells: evidence of a role for extracellular leukotriene A4. *Proc. Natl. Acad. Sci. USA* **82**, 6632–6636 (1985).
23. Fiore, S. & Serhan, C.N. Formation of lipoxins and leukotrienes during receptor-mediated interactions of human platelets and recombinant human granulocyte/macrophage colony-stimulating factor-primed neutrophils. *J. Exp. Med.* **172**, 1451–1457 (1990).
24. Ford-Hutchinson, A.W. Leukotriene B4 in inflammation. *Crit. Rev. Immunol.* **10**, 1–12 (1990).
25. Samuelsson, B. Leukotrienes: mediators of immediate hypersensitivity reactions and inflammation. *Science* **220**, 568–575 (1983).
26. Burke, J.A., Levi, R., Guo, Z.G. & Corey, E.J. Leukotrienes C4, D4 and E4: effects on human and guinea-pig cardiac preparations in vitro. *J. Pharmacol. Exp. Ther.* **221**, 235–241 (1982).
27. Roth, D.M. & Lefer, A.M. Studies on the mechanism of leukotriene induced coronary artery constriction. *Prostaglandins* **26**, 573–581 (1983).
28. Wargovich, T., Mehta, J., Nichols, W.W., Pepine, C.J. & Conti, C.R. Reduction in blood flow in normal and narrowed coronary arteries of dogs by leukotriene C4. *J. Am. Coll. Cardiol.* **6**, 1047–1051 (1985).
29. Falk, E., Shah, P.K. & Fuster, V. Coronary plaque disruption. *Circulation* **92**, 657–671 (1995).
30. Dwyer, J.H. *et al.* Arachidonate 5-Lipoxygenase Promoter Genotype, Dietary Arachidonic Acid, and Atherosclerosis. *N. Engl. J. Med.* **350**, 29–37 (2004).
31. Aiello, R.J. *et al.* Leukotriene B4 receptor antagonism reduces monocytic foam cells in mice. *Arterioscler. Thromb. Vasc. Biol.* **22**, 443–449 (2002).
32. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat. Genet.* **35**, 131–138 (2003).
33. Gretarsdottir, S. *et al.* Localization of a susceptibility gene for common forms of stroke to 5q12. *Am. J. Hum. Genet.* **70**, 593–603 (2002).
34. Gudmundsson, G. *et al.* Localization of a gene for peripheral arterial occlusive disease to chromosome 1p31. *Am. J. Hum. Genet.* **70**, 586–592 (2002).
35. Gulcher, J.R., Kristjansson, K., Gudbjartsson, H. & Stefansson, K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur. J. Hum. Genet.* **8**, 739–742 (2000).
36. Kong, A. & Cox, N.J. Allele-sharing models: LOD scores and accurate linkage tests. *Am. J. Hum. Genet.* **61**, 1179–1188 (1997).
37. Gudbjartsson, D.F., Jonasson, K., Frigge, M.L. & Kong, A. Allegro, a new computer program for multipoint linkage analysis. *Nat. Genet.* **25**, 12–13 (2000).
38. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nat. Genet.* **31**, 241–247 (2002).
39. Whittemore, A.S. & Halpern, J. A class of tests for linkage using affected pedigree members. *Biometrics* **50**, 118–127 (1994).
40. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* **58**, 1347–1363 (1996).
41. Nicolae, D. *Allele Sharing Models in Gene Mapping: A Likelihood Approach* (University of Chicago, 1999).
42. Dempster, A., Laird, N.M. & Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* **39**, 1–38 (1977).
43. Terwilliger, J.D. & Ott, J. A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* **42**, 337–346 (1992).
44. Falk, C.T. & Rubinstein, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann. Hum. Genet.* **51** Pt 3, 227–233 (1987).
45. Lewontin, R.C. The interaction of selection and linkage. ii. Optimum models. *Genetics* **50**, 757–782 (1964).
46. Hill, W.G. & Robertson, A. The effects of inbreeding at loci with heterozygote advantage. *Genetics* **60**, 615–628 (1968).
47. Chen, X., Zehnbauer, B., Gnirke, A. & Kwok, P.Y. Fluorescence energy transfer detection as a homogeneous DNA diagnostic method. *Proc. Natl. Acad. Sci. USA* **94**, 10756–10761 (1997).
48. Steeds, R., Adams, M., Smith, P., Channer, K. & Samani, N.J. Distribution of tissue plasminogen activator insertion/deletion polymorphism in myocardial infarction and control subjects. *Thromb. Haemost.* **79**, 980–984 (1998).
49. Brouilette, S., Singh, R.K., Thompson, J.R., Goodall, A.H. & Samani, N.J. White cell telomere length and risk of premature myocardial infarction. *Arterioscler. Thromb. Vasc. Biol.* **23**, 842–846 (2003).
50. Nomenclature and criteria for diagnosis of ischemic heart disease. Report of the Joint International Society and Federation of Cardiology/World Health Organization task force on standardization of clinical nomenclature. *Circulation* **59**, 607–609 (1979).

Am. J. Hum. Genet. 70:593–603, 2002

Localization of a Susceptibility Gene for Common Forms of Stroke to 5q12

Solveig Gretarsdottir,[1] Sigurlaug Sveinbjörnsdottir,[2] Hjörtur H. Jonsson,[1] Finnbogi Jakobsson,[2] Elisabet Einarsdottir,[1] Uggi Agnarsson,[3] Dana Shkolny,[1] Gisli Einarsson,[2] Herdis M. Gudjonsdottir,[1] Einar M. Valdimarsson,[2] Olafur B. Einarsson,[1] Gudmundur Thorgeirsson,[3] Radinka Hadzic,[1] Sif Jonsdottir,[1] Sigridur Th. Reynisdottir,[1] Sigrun M. Bjarnadottir,[1] Thorunn Gudmundsdottir,[1] Gudrun J. Gudlaugsdottir,[3] Ramanjit Gill,[4] Klaus Lindpaintner,[4] Jesus Sainz,[1] Helgi H. Hannesson,[1] Gunnar Th. Sigurdsson,[1] Michael L. Frigge,[1] Augustine Kong,[1] Vilmundur Gudnason,[3] Kari Stefansson,[1] and Jeffrey R. Gulcher[1]

[1]deCODE Genetics, [2]National University Hospital, and [3]Icelandic Heart Association Heart Preventive Clinic, Reykjavik; and [4]F. Hoffmann–La Roche, Basel, Switzerland

Stroke is one of the most complex diseases, with several subtypes, as well as secondary risk factors, such as hypertension, hyperlipidemia, and diabetes, which, in turn, have genetic and environmental risk factors of their own. Here, we report the results of a genomewide search for susceptibility genes for the common forms of stroke. We cross-matched a population-based list of patients with stroke in Iceland with an extensive computerized genealogy database clustering 476 patients with stroke within 179 extended pedigrees. Linkage to 5q12 was detected, and the LOD score at this locus meets the criteria for genomewide significance (multipoint allele-sharing LOD score of 4.40, $P = 3.9 \times 10^{-6}$). A 20-cM region on 5q was physically and genetically mapped to obtain accurate marker order and intermarker distances. This locus on 5q12, which we have designated as *"STRK1,"* does not correspond to known susceptibility loci for stroke or for its risk factors and represents the first mapping of a locus for common stroke.

Introduction

Stroke is a major health problem in western societies. It is the most common cause of disability, the second-most-common cause of dementia, and the third-most-common cause of death (Bonita 1992). Since it is more common in the elderly, the public health impact of stroke will increase in the next decades with growing life expectancy. Approximately one in four men and one in five women aged 45 years will have a stroke if they live to their 85th year (Bonita 1992). Strategies to diminish the impact of stroke include prevention and treatment with thrombolytic and, possibly, neuroprotective agents. The success of preventive measures will depend on the identification of risk factors and means to modulate their impact.

The clinical phenotype of stroke is complex but can be broadly divided into ischemic and hemorrhagic strokes. The majority (80%–90%) of strokes are ischemic—that is, they are caused by obstruction of blood flow through extra- or intracranial vessels (Mohr et al. 1978; Caplan 2000). The remainder (10%–20%) are hemorrhagic—that is, they result from ruptures of intracranial vessels. Ischemic stroke can be further subdivided into large-vessel occlusive disease, small-vessel occlusive disease, and cardiogenic stroke. For the purposes of this study, we have included transient ischemic attack (TIA) as a biological equivalent of ischemic stroke, even though TIA is not defined as a stroke (because the signs and symptoms, which are the same as those for stroke, last for a short period of time [i.e., <24 h; usually 5–10 min]). This is done because the same pathophysiological mechanisms are considered responsible for TIA and ischemic stroke (Caplan 2000).

The predominant risk factor for all types of stroke is hypertension (Thompson and Furlan 1997; Agnarsson et al. 1999). Hypertension is in itself a complex disease, as are the other known risk factors, diabetes and hyperlipidemia. In addition, there are environmental risk factors, such as smoking. Stroke is therefore considered to be a highly complex disease consisting of a group of heterogeneous disorders with multiple risk factors, both genetic and environmental.

The identification of genetic determinants of common diseases, such as stroke, that may result from the interplay of multiple genes and interactions between genes

Received May 22, 2001; accepted for publication December 12, 2001; electronically published February 6, 2002.

Address for correspondence and reprints: Dr. Kari Stefansson, deCODE Genetics, Sturlugata 8, IS 101-Reykjavik, Iceland. E-mail: kstefans@decode.is; or Dr. Jeffrey Gulcher, deCODE Genetics, Sturlugata 8, IS 101-Reykjavik, Iceland. E-mail: jgulcher@decode.is

0002-9297/2002/7003-0005$15.00

and environment has proven to be a difficult task. Studies of the genetic contribution to stroke have mainly focused either on rare Mendelian diseases in which stroke is a part of the phenotype or on finding association between stroke and possible candidate genes, such as genes contributing to hypertension or lipid metabolism. Several genes have been identified that play roles in the pathogenesis of rare stroke syndromes, such as *Notch3,* in cerebral autosomal dominant arteriopathy with subcortical infarctions and leukoencephalopathy (Tournier-Lasserve et al. 1993; Joutel et al. 1996); *Cystatin C,* in the Icelandic type of hereditary cerebral hemorrhage with amyloidosis (Palsdottir et al. 1988); *APP,* in the Dutch type of hereditary cerebral hemorrhage (Levy et al. 1990); and *KRIT1,* in hereditary cavernous angioma (Gunel et al. 1995; Laberge-le Couteulx et al. 1999; Sahoo et al. 1999).

To our knowledge, no genomewide search for stroke genes in patients with the common forms of stroke has ever been reported. Here we report the results of a genomewide search for susceptibility genes in common stroke by use of a broad but rigorous definition of the phenotype, including hemorrhagic stroke, ischemic stroke, and TIA. The result of this is the mapping of the first major locus reported in common stroke.

Subjects and Methods

Patients

An encrypted population-based list that contained 2,000 living Icelandic patients with stroke and was based on hospital *International Classification of Diseases, Ninth Revision* codes covering the period of time from 1993 to 1997 was run through our computerized genealogy database (Gulcher and Stefansson 1998; Gulcher et al. 2000), which covers the whole Icelandic nation. We excluded patients with subarachnoid hemorrhage or the Icelandic type of hereditary cerebral hemorrhage with amyloidosis. The distribution of stroke types in our study is similar to that reported in other white populations, with ~67% having ischemic strokes, 27% having TIAs, and 6% having hemorrhagic strokes (Caplan 2000). All patients underwent computerized tomography of the head, and the majority of patients underwent Doppler ultrasound of carotid arteries and echocardiography; Holter monitoring was frequently used.

We collected patients with stroke and/or TIAs by use of the criterion that the relationship between each patient and at least one additional patient was characterized by no more than six meiotic events (six meiotic events separate second cousins). Participating patients were more carefully phenotyped by the clinicians before their genotypes were generated. Patients with ischemic stroke and TIAs were classified according to the TOAST (Trial of Org 10172 in Acute Stroke Treatment) subclassification system (Adams et al. 1993). This system includes five categories: (1) large-artery atherosclerosis, (2) cardioembolism, (3) small-artery occlusion (lacune), (4) stroke of other determined etiology, and (5) stroke of undetermined etiology. The diagnoses were based on clinical features and on data from ancillary diagnostic studies. Patients classified as having large-artery atherosclerosis had clinical and brain-imaging findings of cerebral cortical dysfunction and either significant (>70%) stenosis (this is a stricter criteria than that used in TOAST, in which 50% stenosis is the cutoff) or occlusion of a major brain artery or branch cortical artery. Potential sources of cardiogenic embolism were excluded. The second category, cardioembolism, included patients with at least one cardiac source for an embolus and with potential large-artery sources of thromobosis and embolism having been eliminated. Patients with small-artery occlusion had one of the traditional clinical lacunar syndromes and no evidence for cerebral cortical dysfunction. A potential cardiac source of embolus and stenosis >70% in an ipsilateral extracranial artery was excluded. The fourth category, acute stroke of other determined etiology, included patients with rare causes of stroke and patients with two or more potential causes of stroke. If the causes of stroke could not be determined despite extensive evaluation, then patients were included in the fifth category, stroke of undetermined etiology. TOAST classification of patients with ischemic stroke and TIA whom we studied is presented in table 1. Apart from the proportion of large-vessel disease, which is lower in the population that we studied, the subtype distribution is similar to those reported in other studies (e.g., Caplan 2000). This is very likely due to the stricter stenosis criterion that we used to classify large-vessel disease.

The present study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. Informed consent was obtained from all patients and their relatives whose DNA samples were used in the linkage analysis.

Genomewide Scan

A genomewide scan was performed on 476 patients and 438 of their relatives, by use of our framework marker set of 1,000 microsatellite markers. We have developed a microsatellite screening set that is based, in part, on the ABI Linkage Marker (version 2) screening set and the ABI Linkage Marker (version 2) intercalating set, in combination with 500 custom-made markers. All markers were extensively tested for robustness, ease of scoring, and efficiency in 4× multiplex PCRs. In our framework marker set, the average spacing between

markers was ~4 cM, with no gaps >10 cM. Marker positions were obtained from the Marshfield map (Center for Medical Genetics, Marshfield Medical Research Foundation), except for the region containing a three-marker putative inversion on chromosome 8 (Jonsdottir et al. 2000; Giglio et al. 2001; Yu et al. 2001). The PCR amplifications were prepared, run, and pooled on Perkin Elmer/Applied Biosystems 877 Integrated Catalyst Thermocyclers with a similar protocol for each marker. The reaction volume was 5 μl, and, for each PCR, 20 ng of genomic DNA was amplified in the presence of 2 pmol of each primer, 0.25 U AmpliTaq Gold, 0.2 mM dNTPs, and 2.5 mM $MgCl_2$ (buffer was supplied by manufacturer). Cycling conditions were 95°C for 10 min, followed by 37 cycles at 94°C for 15 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min. The PCR products were supplemented with the internal size standard, and the pools were separated and detected on an Applied Biosystems model 377 Sequencer by use of Genescan version 3.0 peak-calling software. Alleles were automatically called with the TrueAllele program (Cybergenetics), and the DecodeGT program (deCODE Genetics) was used both to fractionate according to quality and to edit the called genotypes (Palsson et al. 1999). At least 180 Icelandic controls were genotyped for each marker to derive allele frequencies.

Statistical Methods for Linkage Analysis

In our analyses, we used multipoint, affected-only allele-sharing methods to assess the evidence for linkage. All results, including LOD and nonparametric linkage (NPL) scores, were obtained using the program Allegro (Gudbjartsson et al. 2000). We used the S_{pairs} scoring function (Whittemore and Halpern 1994; Kruglyak et al. 1996) and the exponential allele-sharing model (Kong and Cox 1997) to generate the relevant statistics with 1 df. When combining the family scores to obtain an overall score, instead of weighting the families equally (the default of Genehunter [Kruglyak et al. 1996]) or weighting the affected pairs equally, we used a weighting scheme that is halfway between the two in the log scale; the family weights that we used are the geometric means of the weights of the two schemes. Although not identical, this weighting scheme tends to yield results that are similar to those proposed by Weeks and Lange (1988) as an extension of a weighting scheme by Hodge (1984) that was designed for sibships. We computed the *P* value two different ways and here report the less significant result. The first *P* value was computed on the basis of large sample theory; the distribution of $Z_{lr} = \sqrt{2\,[\log_e(10)\mathrm{LOD}]}$ approximates a standard normal random variable under the null hypothesis of no linkage (Kong and Cox 1997). Because the normal approximation may not work well in some small-sample situations, we computed a second *P* value by comparing the observed LOD score with its complete data-sampling distribution under the null hypothesis (Gudbjartsson et al. 2000). When a data set consists of more than a few families, as is the case here, these two *P* values tend to be very similar. To ensure that the result was a true reflection of the information contained in the material, we considered a linkage result significant not only if the *P* value was $< 2 \times 10^{-5}$ (Lander and Kruglyak 1995) but also if the information content in the region was ≥85%. For the families in the present study, an information content of 85% corresponded to a marker density of approximately one marker per centimorgan. The information measure we used has been defined elsewhere (Nicolae 1999) and has been implemented in Allegro. This measure is closely related to a classical measure of information (Dempster et al. 1977), which has the property that it is between zero, if the marker genotypes are completely uninformative, and one, if the genotypes determine the exact amount of allele sharing by descent among the affected relatives.

After obtaining a significant allele-sharing LOD score,

Table 1

Subclassification of Patients with Stroke

Subtype	% Affected among All Patients (n = 476)	% Affected among Patients in Families with NPL >1 (n = 120)
Hemorrhagic	5	6
Ischemic:		
Large vessel[a]	13	13
Small vessel	16	13
Cardioembolic	23	28
Other cause	4	5
More than one subtype or unknown cause	39	35

[a] The definition of ischemic large-vessel disease that we used is stricter than that usually used in TOAST (see "Subject and Methods" section).

we attempted to understand the contribution of this susceptibility locus by fitting a range of parametric models to the data. Even when fitting parametric models, we performed affected-only analyses, in the sense that an individual is classified as either affected or as having unknown disease status. As a consequence, only ratios of penetrances are relevant. We fitted a range of single-locus dominant, additive, and multiplicative models (Risch 1990). With a complex disease such as stroke, none of these simple models are likely to be exactly true, and the effect of a gene and its variants can only be reliably determined after the at-risk variant (or variants) is identified. However, by the calculation of the corresponding contribution to the sibling recurrence-risk ratio, the fitted parametric models do provide some rough idea of how much the gene is contributing to the familial clustering of the disease.

We investigated the contributions, to the identified locus, of several subtypes of and risk factors for stroke. To do this, we utilized the complete family set and considered as affected only the patients with a particular subtype of or risk factor for stroke. In one particular case, to assess whether the LOD-score increase resulting from the subtraction of the 22 patients with hemorrhagic stroke would be likely to occur by chance, we selected 1,000 random sets of 22 patients whose status we then changed to unknown in an analysis. The *P* value we present is the fraction of the 1,000 simulations that produced, at the peak locus, a LOD-score increase that was equal to or greater than that which we observed by changing the affection status of the patients with hemorrhagic stroke to unknown.

Physical Mapping

To obtain correct marker order and sequence-ready contigs, we physically mapped a 20-cM region, on 5q, that was indicated in the genomewide scan. BAC contigs were generated by a method that combines the results of coincident primer-hybridization experiments with the mining of publicly available sequences. RPCI-11 human male BAC library segments 1 and 2 (Pieter de Jong, Children's Hospital Oakland Research Institute), containing ~200,000 clones with a 12× coverage of the genome, were arrayed using a 6 × 6 double-offset pattern on 23-cm × 23-cm membranes. Initially, hybridizations were performed with markers that were expected, on the basis of their locations in the Weizmann Institute of Science Unified Database for Human Genome Mapping, to be in the region of interest. We used 150 markers in the region (i.e., 31 polymorphic markers used in linkage and 120 markers generated from sequence-tagged sites), which were separated by, on average, 130 kb. The selected markers were used to generate two [^{32}P]-labeled probes: F, which contained the pooled forward primers, and R, which contained the pooled reverse primers. The coincident signals in both hybridizations were selected as positive clones. A set of overlapping clones was assembled through a combination of hybridization and BAC-fingerprint walking. Fingerprints of positive clones (FPCs) were analyzed using the FPC database developed at the Wellcome Trust Sanger Institute. Data from FPC contigs prebuilt with a cutoff of $3e^{-12}$ and from sequence data mining were integrated with the hybridization results. BACs in the region detected by data mining and hybridization were rearrayed. Small membranes (8 cm × 12 cm) were arrayed in 6 × 6 double-offset pattern and were individually hybridized with the markers of interest. A visual map was generated by combining the hybridization, fingerprinting, and sequence data. A total of 137 new markers were generated from BAC end sequences, and the process was repeated until the majority of gaps were closed. Estimates of contig lengths and of the distance between markers assigned to them were based on the FPC program.

Genetic Mapping

High-resolution genetic mapping was used to order contigs obtained by physical mapping and to determine their orientation. In addition to correct marker order, the high-resolution genetic map also provided better estimates of intermarker distances, both of which are important for an accurate linkage analysis (Halpern and Whittemore 1999; Daw et al. 2000). Data from 112 Icelandic nuclear families (sibships with genotypes for two to seven siblings and both parents) were analyzed together with the genotypes for nuclear families available within the stroke pedigrees. For the purpose of genetic mapping, the 112 families alone provide 588 meiotic events, and the inclusion of the data from the families with stroke yielded a map based on substantially more than 1,000 meiotic events. By comparison, the Marshfield genetic map (Center for Medical Genetics, Marshfield Medical Research Foundation) was constructed on the basis of 182 meiotic events. The large number of meiotic events within our families provides the ability to map markers to a resolution $\leq$1.0 cM. In evaluating one order of the markers versus another, by modifying the Allegro program, we computed the number of obligate crossovers for each order, and the order associated with a lower number of crossovers was preferred (Thompson 1987). Given an order, genetic distances between markers were estimated by implementing the expectation-maximization algorithm (Dempster et al. 1977) within the Allegro program. Combining the information from genetic mapping with the physical map resulted in a highly reliable order of markers and intermarker distances within this 20-cM region.

Results

We collected samples from a total of 476 patients, each of whom is related to (within and including six meiotic events) at least one other patient. Patients with hemorrhagic stroke clustered in families with ischemic stroke and TIA, and there were no families with a striking preponderance of either hemorrhagic stroke or further subtypes of ischemic stroke. Given this observation, we decided to study stroke as a broadly defined phenotype. The genome scan was performed with the 476 patients clustered into 179 families. The mean separation of affected pairs was 4.8 meiotic events. Figure 1, which displays four of the families, shows that several stroke subtypes, including hemorrhagic stroke, are found mixed together within the same pedigrees.

Figure 2 presents the allele-sharing LOD scores from the genome scan by use of the framework map. Three regions achieved a LOD score >1.0. Two of these regions were on 5q: one peak at ~69 cM, with a LOD score of 2.00, and a second peak at 99 cM, with a LOD score of 1.14. The third region is on 14q, at 55 cM, with a LOD score of 1.24.

The information for analysis of linkage at the 5q locus was increased by genotyping 45 additional markers over a 45-cM segment that contains both of the regions on 5q (fig. 3). Although the LOD score at the second peak decreased slightly, to ~1.05, the LOD score at the first peak increased to 3.39. However, close inspection of our results suggested not only that the Marshfield genetic map (Center for Medical Genetics, Marshfield Medical Research Foundation) lacks resolution (i.e., many markers were assigned to the same location) but also that there may be some errors in their order. When we followed the marker order of the Marshfield map and used the Allegro program and our data to estimate

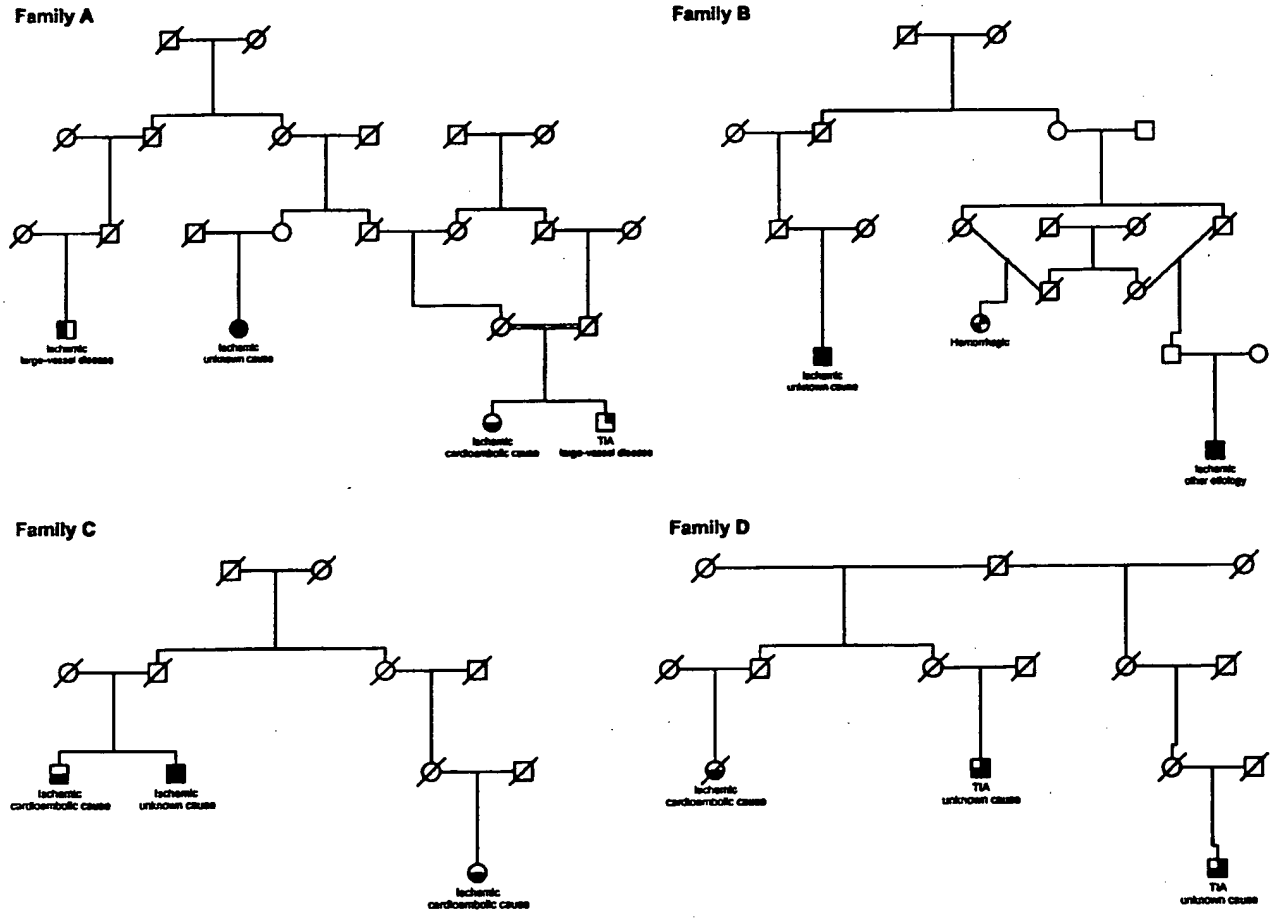


Figure 1 Four families with positive LOD scores. These families include patients with a variety of stroke subphenotypes, as defined by TOAST subclassification (as labeled underneath shaded symbols). Squares and circles represent males and females, respectively; slash marks through symbols indicate individuals who are deceased. Some sex indicators in the two upper generations of the pedigrees have been altered, and unaffected siblings of patients are not displayed, to protect the confidentiality of these families.

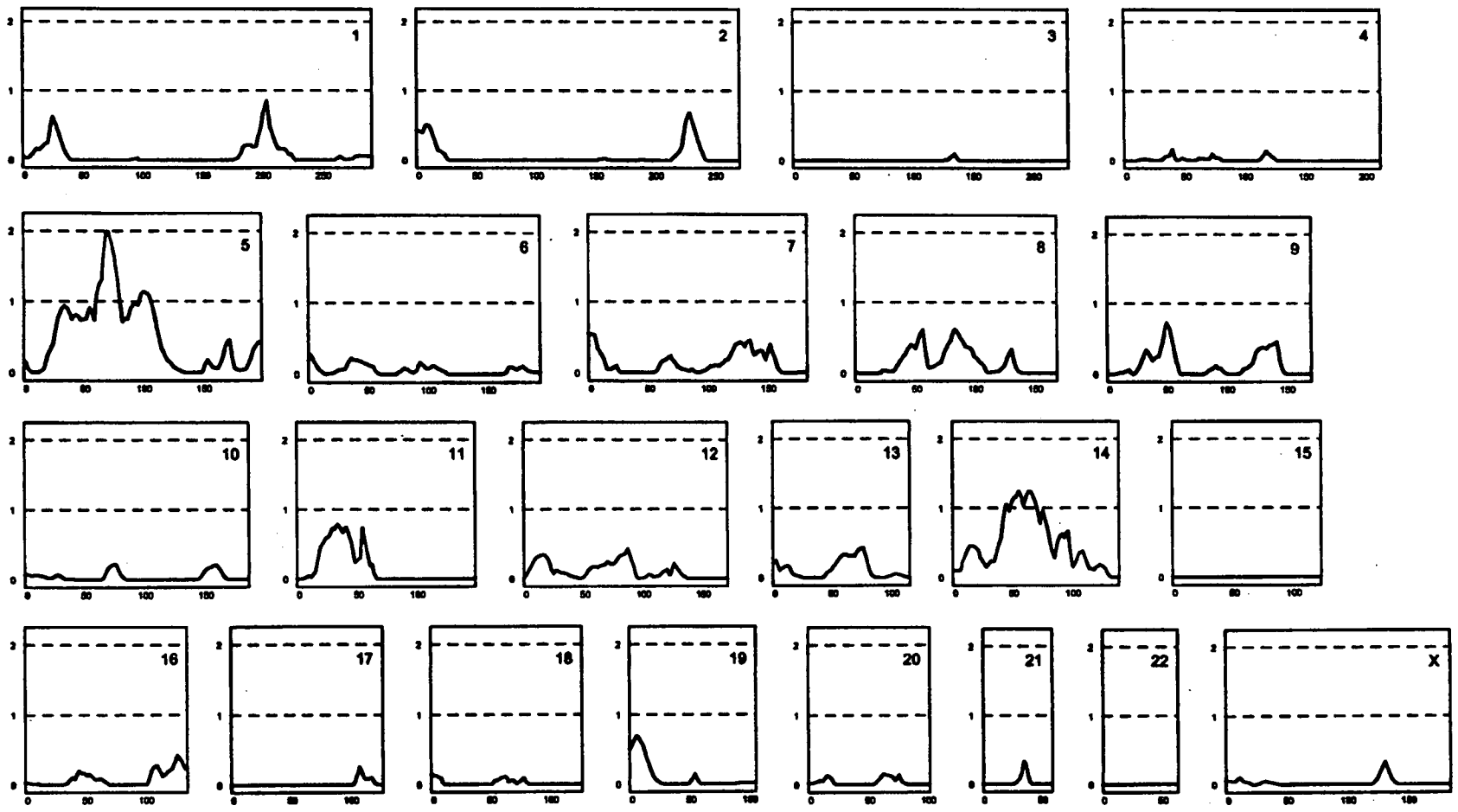


Figure 2 Genomewide scan for stroke-susceptibility loci. A framework map of ~1,000 markers was used. Each box represents a chromosome (indicated in the upper-right-hand corner of each box). The *X*-axis gives the genetic distance (in cM), and the *Y*-axis gives the LOD scores.

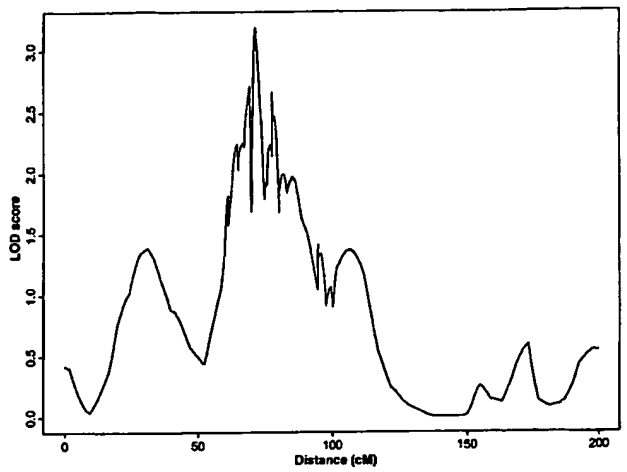


Figure 3 Dense mapping of stroke locus on chromosome 5, with 45 additional markers across the two peaks on 5q. The analysis used the marker order of the Marshfield map (Center for Medical Genetics, Marshfield Medical Research Foundation). The *X*-axis gives the genetic distance (in cM) along the chromosome, and the *Y*-axis gives the LOD score.

the genetic distances between markers, we found that our estimate of the genetic length of the region was substantially longer than that given in the Marshfield map. This indicates a problem with marker order because, in general, incorrect marker order leads to an increased number of apparent crossovers and increases the apparent genetic length. We improved the marker order and intermarker distances by constructing high-density physical and genetic maps over a 20-cM region between D5S474 and D5S2046 (fig. 4). It is worth noting that, although our final order and intermarker distances deviate from those of Marshfield, the overall genetic length for the region is similar.

Linkage analysis with genotypes from the higher-density markers by use of our marker order resulted in a LOD score of 4.40 ($P = 3.9 \times 10^{-6}$) on 5q12 at D5S2080. We designate this locus as "*STRK1*." With the addition of these extra markers, we were able to narrow the most promising region for the harboring of a stroke-susceptibility gene to a segment <6 cM, from D5S1474 to D5S398, as defined by a decrease of 1 in LOD score. Analyses with marker orders based on publicly available marker maps yielded lower LOD scores, 2.78–3.94, thereby highlighting the importance of accurate marker order when using multipoint analysis (fig. 5).

In an attempt to understand the contribution of this susceptibility locus to stroke, we fitted a range of parametric models to the data. The highest LOD score, 4.70, was obtained from a multiplicative model under the assumptions that the at-risk allele frequency was 27% and that there was a fivefold increase in risk for every at-risk allele carried. Under this model, the contribution of this gene to the sibling recurrence-risk ratio was 1.86. Seventy-five of the 179 family clusters yielded a positive LOD score; of these, 55 had LOD scores >0.1, and 5 had LOD scores >0.4. The four families displayed in figure 1 (i.e., families A–D) yielded LOD scores of 0.39, 0.40, 0.47, and 0.48, respectively. These results support the existence of a major stroke-susceptibility gene in this region.

The fractions of all patients in the study who have the various subtypes of stroke are listed in table 1. The fractions are also listed for those families with an NPL score >1 (within these families, there is more sharing among affected members of genetic material at the locus than was expected owing simply to their relationship). The families with more excess sharing at the locus do not show any substantial difference in phenotype pattern from the entire family set. Similar fractions are presented for the risk factors for stroke in each of the two family sets (table 2). Again, no substantial shift in the prevalence of the risk factors is obvious. To assess more directly the contribution of the various subtypes and risk factors to the peak locus, linkage runs were conducted in which only patients with the particular subtype or risk factor were considered as affected—that is, all other patients had their affection status changed to unknown for these runs. In each of these runs, the LOD scores were positive but were smaller than those in the run including all patients. These decreases in LOD score were consistent with the loss of power in the smaller sample sizes. We also conducted a run in which only patients with ischemic stroke were considered as affected. This run, which excluded the 22 patients with hemorrhagic stroke, had an increase in LOD score. The allele-sharing LOD score for this run increased to 4.86 at D5S2080. Although this 0.46 increase in LOD score suggests that *STRK1* is involved primarily in ischemic strokes and TIAs, the increase itself is not statistically significant, on the basis of simulations (one-sided $P = .09$). In summary, these results are consistent with a susceptibility gene at this locus that contributes to a broad spectrum of patients with stroke, the possible exception being patients with hemorrhagic stroke.

Discussion

In this study, we have successfully mapped a major locus for one of the most complex diseases known, by combining genealogy, a comprehensive population-based list of patients with broadly defined stroke, and allele-sharing methods. In any linkage or association study that uses multipoint marker analysis, a correct marker order and precise intermarker distances are important. Otherwise,

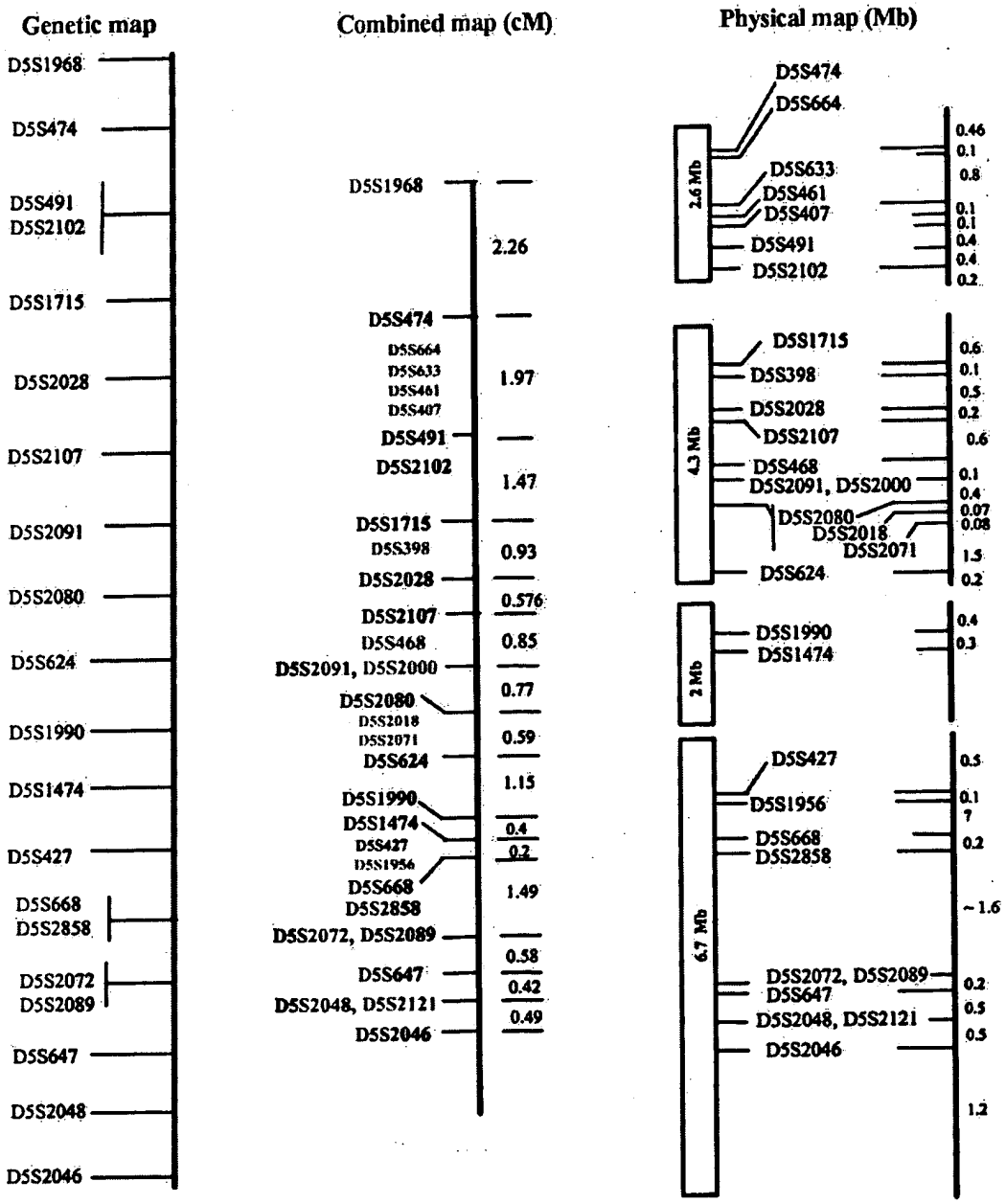


Figure 4 Genetic, physical, and combined maps for *STRK1*, on 5q, from D5S1968 to D5S2046. Markers assigned in both the genetic map and the physical map are displayed in black, markers derived only from physical-map information are displayed in red, and markers derived only from genetic-map information are displayed in blue. Marker distances (in cM) in the combined map were constructed by applying the estimation-maximization algorithm to the final marker order; marker distances (in Mb) in the physical map are estimations from the FPC program.

the apparent increase in information content is neutralized or reduced by the resulting misinformation. We found that a direct application of most public genetic and physical maps, which may have numerous inaccuracies or ambiguities, have a negative impact on the LOD score of this locus for stroke. While our work was in progress, an assembly of the current draft of the human genome—the University of California–Santa Cruz (UCSC) Human Genome Project Working Draft—was made available (Lander et al. 2001). This assembly merges together

overlapping fragments and orders and orients nonoverlapping fragments on the basis of mRNA, EST, paired plasmid reads, and other information. In the April 2001 freeze from UCSC (for which data was released in June 2001), 30 of our 31 linkage markers mapped to two contigs (the remaining marker was not mapped in this freeze). The marker order within the contigs was in agreement with our order, with the exception of two markers, D5S2858 and D5S668. However, in the latest release, in October 2001 (August 2001 freeze), several changes have occurred. Whereas the order of the two markers (D5S2858 and D5S668) has been changed and is now consistent with our order, two other pieces—one involving D5S2028–D5S2080 (>1 Mb) and the other involving D5S427 and D5S1956 (~200 kb)—are flipped and thus are inconsistent with what we believe to be the correct order. This indicates that there is still substantial uncertainty in the assembly of the public human sequence.

The types of stroke that are presented in this article do not reflect a rare stroke form or a form specific to Iceland. Rather, the diverse stroke phenotypes in Icelanders, as well as known risk factors for stroke, are similar to those of most other white populations (Sveinbjörnsdottir et al. 1998; Valdimarsson et al. 1998; Eliasson et al. 1999).

The known genetic factors contributing to common stroke may act indirectly, by increasing the risk of some predisposing conditions, such as diabetes, hyperlipidemias, and/or hypertension. It is also possible that there are genetic factors for stroke that do not influence susceptibilities to the known risk factors, as has been suggested by epidemiological studies for myocardial infarction (Shea et al. 1984; Friedlander et al. 1985; Myers et al. 1990). Epidemiological studies of the common forms of stroke have given conflicting results in regard to the role of family history. Some studies have shown that parental history predicts the risk of stroke

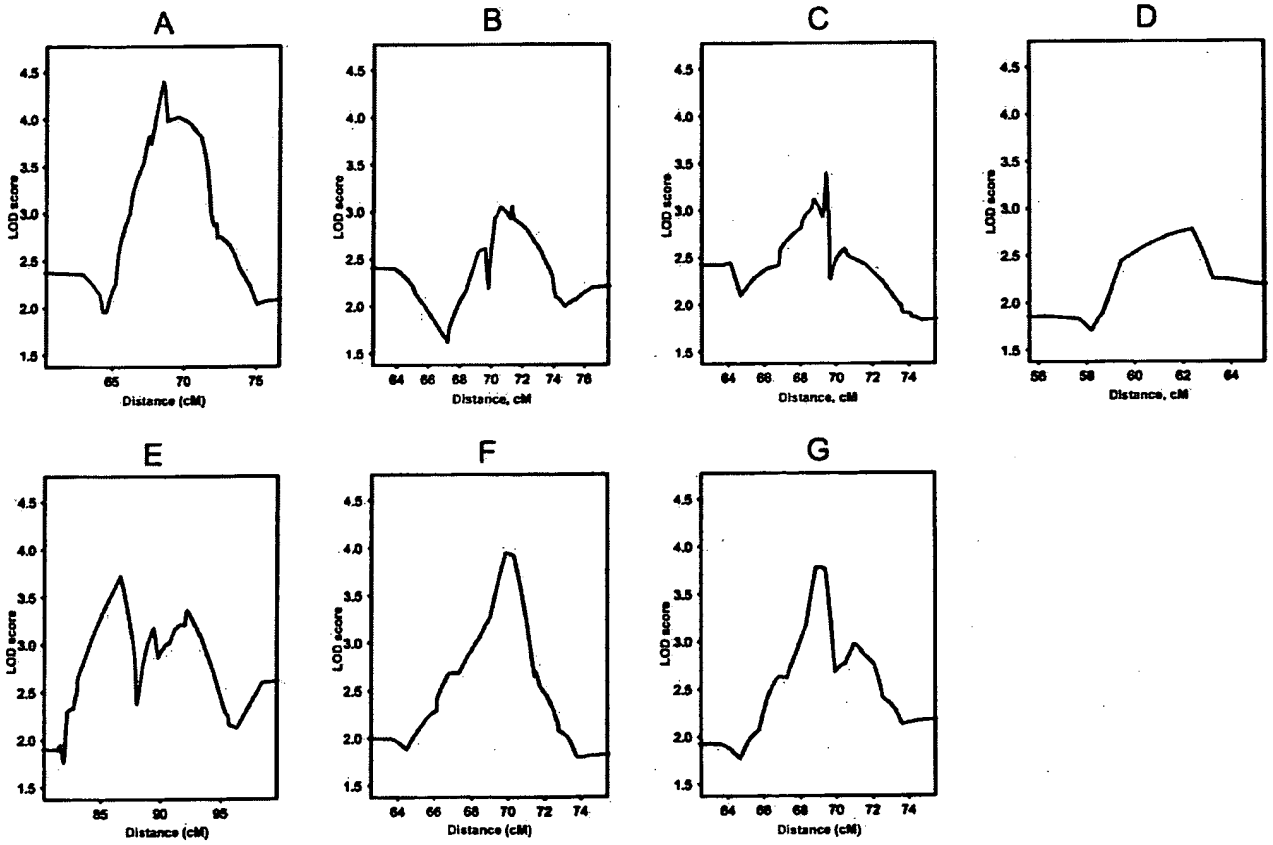


Figure 5 Linkage analyses with marker orders from our combined map (*A;* see fig. 4) and maps from Généthon (*B*), the Center for Medical Genetics, Marshfield Medical Research Foundation (*C;* where the Marshfield map has no resolution, our order was used), Stanford (*D;* radiation-hybrid map), the Weizman Institute Unified Database for Human Genome Mapping (*E*), Whitehead Institute Center for Genome Research (since the Whitehead map only gives order and not distances, it was run both with distances based on application of the estimation-maximization algorithm [*F*] and with equally spaced markers [*G*]).

Table 2

Prevalence of Risk Factors

Risk Factor	% [No.] Affected among	
	All Patients (n = 453)	Patients in Families with NPL >1 (n = 117)
Hypertension[a]	73 [329]	76 [89]
Diabetes[b]	14 [63]	15 [18]
Hypercholesterolemia[c]	24 [111]	21 [25]

Note.—For 23 patients, information on risk factors was unavailable.

[a] If patients (*a*) had measured blood-pressure values of SBP ≥160 mmHg and/or DBP ≥95 mmHg, (*b*) had a history of hypertension, or (*c*) had no history of hypertension but were being treated for hypertension.

[b] If patients (*a*) had nonfasting glucose levels ≥10 mM, (*b*) had a history of diabetes, or (*c*) had no history but were being treated for diabetes.

[c] If patients (*a*) had total cholesterol ≥7 mM or (*b*) were on lipid-lowering medication.

independently from conventional risk factors (Jousilahti et al. 1997; Liao et al. 1997), whereas others have failed to find evidence for such independent factors (Kiely et al. 1993; Lindenstrom et al. 1993; Graffagnino 1994). However, our work describes the first reported genome scan in search of genes that contribute to common forms of stroke. Our data suggest that the locus we have mapped contributes directly to stroke, rather than indirectly through known risk factors for stroke. This suggests that there may be biological pathways independent of the known risk factors that contribute to the pathogenesis of stroke. Regardless of what the mechanism is, the evidence presented supports a major genetic component in the pathogenesis of stroke in Iceland.

Acknowledgments

We thank the participating Icelandic patients with stroke and their families, whose contribution have made this study possible; the nurses at the Icelandic Heart Association, Hólmfrídur B. Thorsteinsdottir, and the members of deCODE's BAC physical-mapping and genotyping facilities, for excellent assistance; and Dr. Thorgeir Thorgeirsson, for valuable comments on the manuscript.

Electronic-Database Information

URLs for data in this article are as follows:

Center for Medical Genetics, Marshfield Medical Research Foundation, http://research.marshfieldclinic.org/genetics/ (for genetic marker map)

Cybergenetics, http://www.cybgen.com/ (for TrueAllele program)

deCODE Genetics, http://www.decode.com/ (for DecodeGT program)

Généthon, http://www.genethon.fr/

UCSC Human Genome Project Working Draft ("Golden Path"), http://genome.ucsc.edu/

Unified Database for Human Genome Mapping, The, http://bioinformatics.weizmann.ac.il/udb/

Wellcome Trust Sanger Institute, The, http://www.sanger.ac.uk/

Whitehead Institute Center for Genome Research, http://www-genome.wi.mit.edu/

References

Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE (1993) Classification of subtype of acute ischemic stroke: definitions for use in a multicenter clinical trial. Stroke 24:35–41

Agnarsson U, Thorgeirsson G, Sigvaldason H, Sigfusson N (1999) Effects of leisure-time physical activity and ventilatory function on risk for stroke in men: the Reykjavik Study. Ann Intern Med 130:987–990

Bonita R (1992) Epidemiology of stroke. Lancet 339:342–344

Caplan LR (2000) Caplan's stroke: a clinical approach, 3d ed. Butterworth-Heinemann, Stoneham, MA, pp 1–556

Daw EW, Thompson EA, Wijsman EM (2000) Bias in multipoint linkage analysis arising from map misspecification. Genet Epidemiol 19:366–380

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 39:1–38

Eliasson JH, Valdimarsson EM, Jakobsson F (1999) Case fatality after acute stroke at the Reykjavik Hospital in 1996–1997. Læknablaðið 85:517–525

Friedlander Y, Kark JD, Stein Y (1985) Family history of myocardial infarction as an independent risk factor for coronary heart disease. Br Heart J 53:382–387

Giglio S, Broman KW, Matsumoto N, Calvari V, Gimelli G, Neumann T, Ohashi H, Voullaire L, Larizza D, Giorda R, Weber JL, Ledbetter DH, Zuffardi O (2001) Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. Am J Hum Genet 68:874–883

Graffagnino C (1994) The importance of family history in cerebrovascular disease. Stroke 25:1599–1603

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25:12–13

Gulcher JR, Kristjansson K, Gudbjartsson H, Stefansson K (2000) Protection of privacy by third-party encryption in genetic research in Iceland. Eur J Hum Genet 8:739–742

Gulcher J, Stefansson K (1998) Population genomics: laying the groundwork for genetic disease modeling and targeting. Clin Chem Lab Med 36:523–527

Gunel M, Awad IA, Anson J, Lifton RP (1995) Mapping a gene causing cerebral cavernous malformation to 7q11.2-q21. Proc Natl Acad Sci USA 92:6620–6624

Halpern J, Whittemore AS (1999) Multipoint linkage analysis: a cautionary note. Hum Hered 49:194–196

Hodge SE (1984) The information contained in multiple sibling pairs. Genet Epidemiol 1:109–122

Jonsdottir GM, Jonasson K, Gudbjartsson D, Kong A (2000)

Construction of a genetic marker map using genotypes from over 7500 Icelanders. Am J Hum Genet 67 Suppl 2:332

Jousilahti P, Rastenyte D, Tuomilehto J, Sarti C, Vartiainen E (1997) Parental history of cardiovascular disease and risk of stroke. A prospective follow-up of 14371 middle-aged men and women in Finland. Stroke 28:1361–1366

Joutel A, Corpechot C, Ducros A, Vahedi K, Chabriat H, Mouton P, Alamowitch S, Domenga V, Cecillion M, Marechal E, Maciazek J, Vayssiere C, Cruad C, Cabanis EA, Ruchoux MM, Wessenbach J, Bach JF, Bousser MG, Tournier-Lasserve E (1996) *Notch3* mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. Nature 383:707–710

Kiely DK, Wolf PA, Cupples LA, Beiser AS, Myers RH (1993) Familial aggregation of stroke: the Framingham Study. Stroke 24:1366–1371

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Laberge-le Couteulx S, Jung HH, Labauge P, Houtteville JP, Lescoat C, Cecillon M, Marechal E, Joutel A, Bach JF, Tournier-Lasserve E (1999) Truncating mutations in *CCM1*, encoding KRIT1, cause hereditary cavernous angiomas. Nat Genet 23:189–193

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Levy E, Carman MD, Fernandez-Madrid IJ, Power MD, Lieberburg I, van Duinen SG, Bots GT, Luyendijk W, Frangione B (1990) Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type. Science 248: 1124–1126

Liao D, Myers R, Hunt S, Shahar E, Paton C, Burke G, Province M, Heiss G (1997) Familial history of stroke and stroke risk: the Family Heart Study. Stroke 28:1908–1912

Lindenstrom E, Boysen G, Nyboe J (1993) Risk factors for stroke in Copenhagen, Denmark. I. Basic demographic and social factors. Neuroepidemiology 12:37–42

Mohr JP, Caplan LR, Melski JW, Goldstein RJ, Duncan GW, Kistler JP, Pessin MS, Bleich HL (1978) The Harvard Cooperative Stroke Registry: a prospective registry. Neurology 28:754–762

Myers RH, Kiely DK, Cupples LA, Kannel WB (1990) Parental history is an independent risk factor for coronary artery disease: the Framingham Study. Am Heart J 120:963–969

Nicolae DL (1999) Allele sharing models in gene mapping: a likelihood approach. PhD thesis, Department of Statistics, University of Chicago, Chicago, p 126

Palsdottir A, Abrahamson M, Thorsteinsson L, Arnason A, Olafsson I, Grubb A, Jensson O (1988) Mutation in *cystatin* C gene causes hereditary brain haemorrhage. Lancet 2:603–604

Palsson B, Palsson F, Perlin M, Gudbjartsson H, Stefansson K, Gulcher J (1999) Using quality measures to facilitate allele calling in high-throughput genotyping. Genome Res 9:1002–1012

Risch N (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46:229–241

Sahoo T, Johnson EW, Thomas JW, Kuehl PM, Jones TL, Dokken CG, Touchman JW, Gallione CJ, Lee-Lin SQ, Kosofsky B, Kurth JH, Louis DN, Mettler G, Morrison L, Gil-Nagel A, Rich SS, Zabramski JM, Boguski MS, Green ED, Marchuk DA (1999) Mutations in the gene encoding KRIT1, a Krev-1/rap1a binding protein, cause cerebral cavernous malformations (*CCM1*). Hum Mol Genet 8:2325–2333

Shea S, Ottman R, Gabrieli C, Stein Z, Nichols A (1984) Family history as an independent risk factor for coronary artery disease. J Am Coll Cardiol 4:793–801

Sveinbjörnsdottir S, Einarsson G, Magnusdottir S, Gudmundsson G, Jonsson JE (1998) Systematic registration of patients with stroke and TIA admitted to the National University Hospital, Reykjavik, Iceland, in 1997. Læknablaðið Suppl 36:54

Thompson DW, Furlan AJ (1997) Clinical epidemiology of stroke. Neurosurg Clin N Am 8:265–269

Thompson EA (1987) Crossover counts and likelihood in multipoint linkage analysis. IMA J Math Appl Med Biol 4:93–108

Tournier-Lasserve E, Joutel A, Melki J, Weissenbach J, Lathrop GM, Chabriat H, Mas JL, et al (1993) Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy maps to chromosome 19q12. Nat Genet 3:256–259

Valdimarsson EM, Sigurdsson G, Jakobsson F (1998) Etiology and treatment of cerebral ischemia at the Department of Neurology and Rehabilitation Medicine at Reykjavik City Hospital. Læknablaðið 84:921–927

Weeks DE, Lange K (1988) The affected-pedigree-member method of linkage analysis. Am J Hum Genet 42:315–326

Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. Biometrics 50:118–127

Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW, Weber JL (2001) Comparison of human genetic and sequence-based physical maps. Nature 409:951–953

Am. J. Hum. Genet. 76:000–000, 2005

Report

Association between the Gene Encoding 5-Lipoxygenase–Activating Protein and Stroke Replicated in a Scottish Population

A. Helgadottir,[1] S. Gretarsdottir,[1] D. St. Clair,[2] A. Manolescu,[1] J. Cheung,[2] G. Thorleifsson,[1] A. Pasdar,[2] S. F. A. Grant,[1] L. J. Whalley,[2] H. Hakonarson,[1] U. Thorsteinsdottir,[1] A. Kong,[1] J. Gulcher[1] K. Stefansson,[1] and M. J. MacLeod[2]

[1]deCODE Genetics, Reykjavik; and [2]Aberdeen Royal Infirmary and University of Aberdeen Medical School, Aberdeen, Scotland

Cardiovascular diseases, including myocardial infarction (MI) and stroke, most often occur on the background of atherosclerosis, a condition attributed to the interactions between multiple genetic and environmental risk factors. We recently reported a linkage and association study of MI and stroke that yielded a genetic variant, HapA, in the gene encoding 5-lipoxygenase–activating protein (*ALOX5AP*), that associates with both diseases in Iceland. We also described another *ALOX5AP* variant, HapB, that associates with MI in England. To further assess the contribution of the *ALOX5AP* variants to cardiovascular diseases in a population outside Iceland, we genotyped seven single-nucleotide polymorphisms that define both HapA and HapB from 450 patients with ischemic stroke and 710 controls from Aberdeenshire, Scotland. The Icelandic at-risk haplotype, HapA, had significantly greater frequency in Scottish patients than in controls. The carrier frequency in patients and controls was 33.4% and 26.4%, respectively, which resulted in a relative risk of 1.36, under the assumption of a multiplicative model ($P = .007$). We did not detect association between HapB and ischemic stroke in the Scottish cohort. However, we observed that HapB was overrepresented in male patients. This replication of haplotype association with stroke in a population outside Iceland further supports a role for *ALOX5AP* in cardiovascular diseases.

Cardiovascular diseases (CVDs), such as coronary heart disease and stroke, are major causes of death and disability in western societies (Aboderin et al. 2002). As a result of the increasing age of the population, the prevalence of CVD is rising worldwide (American Heart Association 2002). CVDs are largely attributed to atherosclerosis, which has various environmental and genetic risk factors. It is a commonly held view that chronic inflammation initiates and promotes the development of atherosclerotic lesions (Lusis 2000; Libby 2002). Large epidemiologic studies have demonstrated correlations between increased production of markers of systemic inflammation and future cardiovascular events, including myocardial infarction (MI) (Ridker et al. 1997, 1998; Danesh et al. 2000) and stroke (Di Napoli et al. 2001), which supports a central role for inflammation in CVD.

We recently published the association of a variant in the gene encoding 5-lipoxygenase–activating protein (*ALOX5AP* [MIM 603700]) with both MI and stroke in an Icelandic population (Helgadottir et al. 2004). *ALOX5AP*, which encodes an important component of the leukotriene pathway, was identified through a genomewide linkage scan conducted on 296 families with MI and subsequent analysis that determined association with markers within the mapped region on chromosome 13q12-13. A haplotype spanning *ALOX5AP*, HapA, defined by four SNPs, was shown to be associated with MI (relative risk = 1.8; $P = .0000023$) and, subsequently, the same variant was found to confer risk of stroke in Iceland (relative risk [RR] = 1.7; $P = .000095$) (Helgadottir et al. 2004). Another SNP-based haplotype within *ALOX5AP*, HapB, showed significant association with MI in British cohorts from Leicester and Sheffield (RR = 2.0; $P = .00037$) (Helgadottir et al. 2004). We further demonstrated that leukotriene B4 (LTB4) synthesis by neutrophils from patients with a history of MI

Received September 28, 2004; accepted for publication December 13, 2004; electronically published January 7, 2005.

Address for correspondence and reprints: Dr. K. Stefansson, deCODE Genetics Inc., Sturlugata 8, 101 Reykjavik, Iceland. E-mail: kstefans@decode.is

0002-9297/2005/7603-00XX$15.00

is greater than the synthesis by those from controls without MI (Helgadottir et al. 2004).

In the present study, we attempted to replicate the association of *ALOX5AP* with stroke in a population outside Iceland. The SNPs defining HapA (*SG13S25, SG13S114, SG13S89,* and *SG13S32*) and HapB (*SG13S377, SG13S114, SG13S41,* and *SG13S35*) were genotyped for 450 Scottish patients who had experienced a stroke and for 710 controls. The patient and control cohorts have been described elsewhere (MacLeod et al. 1999; Meiklejohn et al. 2001; Duthie et al. 2002; Whalley et al. 2004). In brief, 450 patients from northeastern Scotland with CT confirmation of ischemic stroke (including 26 patients with transient ischemic attack [TIA]) were recruited between 1997 and 1999, within 1 wk of admission to the Acute Stroke Unit at Aberdeen Royal Infirmary. Patients were further subclassified in accordance with the TOAST (Trial of Org 10172 in Acute Stroke Treatment) research criteria (Adams et al. 1993). Of the patients, 155 (34.4%) had large-vessel stroke, 96 (21.3%) had cardiogenic stroke, and 109 (24.2%) had small-vessel stroke; for 5 (1.1%) of the patients, stroke with other determined etiology was diagnosed, 7 (1.6%) had more than one etiology, and 78 (17.3%) had unknown cause of stroke despite extensive evaluation. A total of 710 control individuals with no history of stroke or TIA were recruited during follow-up of the 1921 ($n = 227$) and 1936 ($n = 371$) Aberdeen Birth Cohort Studies originally recruited in 1932 and 1947, respectively, as part of the Scottish mental surveys (Deary et al. 2004). A further 112 controls were recruited from local primary-care practices (Meiklejohn et al. 2001). Basic clinical characteristics of patients and control individuals are shown in table 1. Approval for the study was granted by the local research ethics committee, and all study participants gave written informed consent.

The haplotype analysis was performed using the program NEMO (Gretarsdottir et al. 2003). NEMO handles missing genotypes and uncertainty with phase through a likelihood procedure, by use of the expectation-maximization algorithm as a computational tool to estimate haplotype frequencies. Since we were testing only two haplotypes, which had been shown elsewhere to confer risk of MI and stroke in an Icelandic cohort and MI in an English cohort, the reported *P* values are one sided. For the at-risk haplotypes, we calculated RR and population-attributable risk (PAR) under the assumption of a multiplicative model (Falk and Rubinstein 1987; Terwilliger and Ott 1992) in which the risk of the two alleles of haplotypes a person carries multiplies.

The results of the haplotype-association analysis for HapA and HapB are shown in table 2. The haplotype frequencies of HapA in the Scottish populations (patient and control) were higher than in the corresponding Icelandic populations (table 2). As demonstrated in the Icelandic population, the estimated frequency of HapA was significantly greater in Scottish patients who have suffered a stroke than in Scottish controls. The carrier frequency of HapA in Scottish patients and controls was 33.4% and 26.4%, respectively, which resulted in an RR of 1.36 ($P = .007$) and a corresponding PAR of 9.6%. We had previously observed in the Icelandic population a higher frequency of HapA in male than in female patients with either stroke or MI (Helgadottir et al. 2004). This sex difference in the frequency of HapA was not observed in the Scottish population (table 2).

Table 1

Clinical Characteristics of Scottish Patients and Control Individuals

Characteristics	Patients ($n = 450$)	Controls ($n = 710$)
Female:male	42:58	49:51
Age (years)	66.8 ± .6	67.2 ± .4
Hypertension (%)	55.5	23.9
Diabetes (%)	12.6	2.1
Total cholesterol (mmol/liter)	5.65 ± .06	5.64 ± .05

NOTE.—Patients and control individuals were classified as having hypertension and/or diabetes on the basis of previous history or receipt of antihypertensive or antidiabetic therapy. Values with plus-minus symbol (±) are mean ± SE.

We then tested the association of HapB with stroke in the Scottish cohort. HapB has been shown elsewhere to confer risk of MI in an English cohort (Helgadottir et al. 2004). A slight excess of HapB was observed in the patient group (6.8%) compared with controls (5.8%), but it was not significant (table 2). However, sex-specific analysis showed that the frequency of HapB was higher in males with ischemic stroke (9.2%) than in controls, resulting in an RR of 1.65 ($P = .016$). The frequency of HapB in females with ischemic stroke was 3.5%, which was lower but not significantly different from that of controls. The frequencies of HapB in males and females with ischemic stroke differed significantly ($P = .0021$). Interestingly, as shown in table 2, similar trends were observed in our Icelandic cohort; the frequency of HapB was greater in males with ischemic stroke (8.6%) than in females with ischemic stroke (5.8%), although this was not significant ($P = .055$).

To summarize our results, we demonstrate in the present study that HapA, the risk haplotype of *ALOX5AP*, reported elsewhere to confer risk of MI and stroke in an Icelandic cohort, associates with ischemic stroke in a Scottish cohort. HapB, which confers risk of MI in an English cohort, was not associated with ischemic stroke in the Scottish cohort. However, we observed that HapB was overrepresented in male patients.

Historical and archaeological data have suggested a Gaelic ancestry for both Icelanders and Scots. This is

Table 2

Analysis of Association of HapA and HapB with Ischemic Stroke

Location and Study Population (*n*)	HapA			HapB		
	Frequency	RR	*P*	Frequency	RR	*P*
Scotland:						
Controls (710)	.142			.058		
Patients with ischemic stroke (450[a]):	.184	1.36	.007	.068	1.20	NS
Males (253)	.183	1.35	.023	.092	1.65	.016
Females (181)	.179	1.34	.044	.035	.58	NS
Iceland:						
Controls (624)	.095			.067		
Patients with ischemic stroke (632):	.147	1.63	.00013	.073	1.09	NS
Males (335)	.155	1.75	.0002	.086	1.31	NS
Females (297)	.138	1.51	.0079	.058	.86	NS

Note.—Shown are HapA and HapB of *ALOX5AP* and the corresponding number of individuals genotyped, the haplotype frequency in the patient and control cohorts, the RR, and the one-sided *P* values. HapA is defined by the SNPs *SG13S25, SG13S114, SG13S89,* and *SG13S32,* with alleles G, T, G, and A, respectively, and HapB is defined by the SNPs *SG13S377, SG13S114, SG13S41,* and *SG13S35,* with alleles A, A, A, and G, respectively. For SNP genotyping, we used TaqMan assays (Applied Biosystems) or the fluorescent-polarization template-directed dye-terminator incorporation (the SNP-FP-TDI assay), as described elsewhere (Chen et al. 1999). SNP information can be found in the dbSNP database. The DNA used for the SNP genotyping was the product of whole-genome amplification, by use of the GenomiPhi Amplification kit (Amersham), of DNA isolated from the peripheral blood of the Scottish controls and patients with stroke. Data on the Icelandic cohort have been reported elsewhere (Helgadottir et al. 2004). NS = not significant.

[a] Sex unknown for 16 patients.

further supported by recent studies of mtDNA and Y-chromosome diallelic and microsatellite variation in Icelanders, Scandinavians, and Gaels from Ireland and Scotland (Helgason et al. 2000, 2001). Given this common ancestry, it is possible that the two populations share a disease-causing variant and that this variant may reside on the same common haplotype background (HapA). Such a scenario would be consistent with our results; although the estimated RR for HapA in the Scottish cohort is somewhat lower than in the Icelandic cohort, this difference is not statistically significant. Indeed, a similar observation has been made in previous studies of schizophrenia in Iceland and Scotland (Stefansson et al. 2003), in which the same extended haplotype was found to confer risk of schizophrenia in both populations, with comparable frequencies in patient and control groups in the two countries.

The gene *ALOX5AP* encodes the membrane-associated 5-lipoxygenase–activating protein (FLAP), an important mediator of the activity of cellular 5-lipoxygenase (5-LO), which is a key enzyme in the biosynthesis of leukotrienes (Dixon et al. 1990; Miller et al. 1990). Leukotrienes are proinflammatory mediators produced predominantly in inflammatory cells such as polymorphonuclear leukocytes, macrophages, and mast cells. Over the last decade, a number of studies have supported an important role for inflammation in atherosclerosis—from atheroma initiation to promotion of plaque rupture, thereby triggering thrombosis, the main atherosclerotic complication that causes MI and stroke (Libby 2002). The 5-LO pathway could be an important contributor to the pathophysiology of atherosclerosis through the formation of the proinflammatory LTB4 and/or through an increase in vascular permeability caused by cysteinyl leukotrienes. Indeed, we have shown increased production of LTB4 in neutrophils from patients with history of MI, compared with controls without history of MI (Helgadottir et al. 2004). This is further supported by recent human-expression studies (Spanbroek et al. 2003) that show an increased expression of members of the 5-LO pathway, including 5-LO and FLAP, in atherosclerotic lesions at various stages of their development. Moreover, a promoter variant of 5-LO (*ALOX5* [MIM 152390]) has been shown to be associated with increased carotid artery intima-media thickness and with heightened inflammatory biomarkers (Dwyer et al. 2004). In addition, an atherosclerotic mouse model with a heterozygous deficiency of 5-LO shows resistance to atherosclerosis (Mehrabian et al. 2002), and an LTB4 receptor antagonist blocks the development of atherosclerosis in apoE- and LDLR-deficient mice (Aiello et al. 2002; Mehrabian et al. 2002). Together, these studies suggest that chronic upregulation of the leukotriene pathway may be harmful to the vasculature, in terms of atherosclerosis progression and plaque instability.

The precise mechanism by which the *ALOX5AP* variants confer risk of MI and stroke is still unclear. As reported elsewhere, we have not observed SNPs in the coding sequence that led to amino acid substitution (Helgadottir et al. 2004). Therefore, one can speculate that

unidentified variation in regulatory regions of the gene—that affects transcription, splicing, message stability, message transport, or translation efficiency—may underlie the risk conferred by *ALOX5AP*.

The results of the present study show that HapA associates with ischemic stroke in a Scottish population, thereby providing replication of work that showed that the same haplotype confers increased risk of stroke in an Icelandic population. This replication constitutes additional evidence for the role of *ALOX5AP* in the pathogenesis of stroke. Identification of genetic risk factors for the common forms of stroke may facilitate identification of individuals at increased risk and may lead to novel strategies for the prevention and treatment of stroke.

Acknowledgments

The authors thank Thorbjorg Jonsdottir and personnel at the deCODE core facilities for valuable contributions. The authors acknowledge the support of the Chief Scientist's Office, Scottish Executive.

Electronic-Database Information

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/SNP/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for *ALOX5AP* and *ALOX5*)

References

Aboderin I, Kalache A, Ben-Shlomo Y, Lynch JW, Yajnik CS, Kuh D, Yach D (2002) Life course perspectives on coronary heart disease, stroke and diabetes: key issues and implications for policy and research. World Health Organization, Geneva

Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE 3rd (1993) Classification of subtype of acute ischemic stroke: definitions for use in a multicenter clinical trial. Stroke 24:35–41

Aiello RJ, Bourassa PA, Lindsey S, Weng W, Freeman A, Showell HJ (2002) Leukotriene B4 receptor antagonism reduces monocytic foam cells in mice. Arterioscler Thromb Vasc Biol 22:443–449

American Heart Association (2002) Heart disease and stroke statistics: 2003 update, Dallas

Chen X, Levine L, Kwok PY (1999) Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res 9:492–498

Danesh J, Whincup P, Walker M, Lennon L, Thomson A, Appleby P, Gallimore JR, Pepys MB (2000) Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses. BMJ 321:199–204

Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC (2004) The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. J Pers Soc Psychol 86:130–147

Di Napoli M, Papa F, Bocola V (2001) C-reactive protein in ischemic stroke: an independent prognostic factor. Stroke 32: 917–924

Dixon RA, Diehl RE, Opas E, Rands E, Vickers PJ, Evans JF, Gillard JW, Miller DK (1990) Requirement of a 5-lipoxygenase-activating protein for leukotriene synthesis. Nature 343:282–284

Duthie SJ, Whalley LJ, Collins AR, Leaper S, Berger K, Deary IJ (2002) Homocysteine, B vitamin status, and cognitive function in the elderly. Am J Clin Nutr 75:908–913

Dwyer JH, Allayee H, Dwyer KM, Fan J, Wu H, Mar R, Lusis AJ, Mehrabian M (2004) Arachidonate 5-lipoxygenase promoter genotype, dietary arachidonic acid, and atherosclerosis. N Engl J Med 350:29–37

Falk CT, Rubinstein P (1987) Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet 51:227–233

Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, Jonsdottir T, Gudmundsdottir T, et al (2003) The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. Nat Genet 35:131–138

Helgadottir A, Manolescu A, Thorleifsson G, Gretarsdottir S, Jonsdottir H, Thorsteinsdottir U, Samani NJ, et al (2004) The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. Nat Genet 36:233–239

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, Sykes B (2001) mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. Am J Hum Genet 68:723–737

Helgason A, Sigurðardóttir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, Gulcher JR, Ward R, Stefánsson K (2000) Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. Am J Hum Genet 67:697–717

Libby P (2002) Inflammation in atherosclerosis. Nature 420: 868–874

Lusis AJ (2000) Atherosclerosis. Nature 407:233–241

MacLeod MJ, Dahiyat MT, Cumming A, Meiklejohn D, Shaw D, St Clair D (1999) No association between glu/asp polymorphism of *NOS3* gene and ischemic stroke. Neurology 53:418–420

Mehrabian M, Allayee H, Wong J, Shi W, Wang XP, Shaposhnik Z, Funk CD, Lusis AJ, Shih W (2002) Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice. Circ Res 91:120–126

Meiklejohn DJ, Vickers MA, Dijkhuisen R, Greaves M (2001) Plasma homocysteine concentrations in the acute and convalescent periods of atherothrombotic stroke. Stroke 32:57–62

Miller DK, Gillard JW, Vickers PJ, Sadowski S, Léveillé C, Mancini JA, Charleson P, Dixon RAF, Ford-Hutchinson AW, Fortin R, Gautier JY, Rodkey J, Rosen R, Rouzer C, Sigal IS, Strader CD, Evans JF (1990) Identification and isolation of a membrane protein necessary for leukotriene production. Nature 343:278–281

Ridker PM, Buring JE, Shih J, Matias M, Hennekens CH (1998) Prospective study of C-reactive protein and the risk of future

cardiovascular events among apparently healthy women. Circulation 98:731–733

Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH (1997) Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. N Engl J Med 336:973–979

Spanbroek R, Grabner R, Lotzer K, Hildner M, Urbach A, Ruhling K, Moos MP, Kaiser B, Cohnert TU, Wahlers T, Zieske A, Plenz G, Robenek H, Salbach P, Kuhn H, Radmark O, Samuelsson B, Habenicht AJ (2003) Expanding expression of the 5-lipoxygenase pathway within the arterial wall during human atherogenesis. Proc Natl Acad Sci USA 100: 1238–1243

Stefansson H, Sarginson J, Kong A, Yates P, Steinthorsdottir V, Gudfinnsson E, Gunnarsdottir S, Walker N, Petursson H, Crombie C, Ingason A, Gulcher JR, Stefansson K, St Clair D (2003) Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. Am J Hum Genet 72: 83–87

Terwilliger JD, Ott J (1992) A haplotype-based "haplotype relative risk" approach to detecting allelic associations. Hum Hered 42:337–346

Whalley LJ, Fox HC, Wahle KW, Starr JM, Deary IJ (2004) Cognitive aging, childhood intelligence, and the use of food supplements: possible involvement of n-3 fatty acids. Am J Clin Nutr 80:1650–1657

Journal of Thrombosis and Haemostasis, 1: 80–87

ORIGINAL ARTICLE

The epidemiology of venous thromboembolism in Caucasians and African-Americans: the GATE Study[1]

N. F. DOWLING,* H. AUSTIN,† A. DILLEY,* C. WHITSETT,‡ B. L. EVATT* and W. C. HOOPER*

**Haematologic Diseases Branch, Division of AIDS, STD, and TB Laboratory Research, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA; †Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA; and ‡Mount Sinai School of Medicine, New York, NY, USA*

Summary. The aim of this study was to assess, comprehensively, medical and genetic attributes of venous thromboembolism (VTE) in a multiracial American population. The Genetic Attributes and Thrombosis Epidemiology (GATE) study is an ongoing case–control study in Atlanta, Georgia, designed to examine racial differences in VTE etiology and pathogenesis. Between 1998 and 2001, 370 inpatients with confirmed VTE, and 250 control subjects were enrolled. Data collected included blood specimens for DNA and plasma analysis and a medical lifestyle history questionnaire. Comparing VTE cases, cancer, recent surgery, and immobilization were more common in caucasian cases, while hypertension, diabetes, and kidney disease were more prevalent in African-American cases. Family history of VTE was reported with equal frequency by cases of both races (28–29%). Race-adjusted odds ratios for the associations of factor V Leiden and prothrombin G20210A mutations were 3.1 (1.5, 6.7) and 1.9 (0.8, 4.4), respectively. Using a larger external comparison group, the odds ratio for the prothrombin mutation among Caucasians was a statistically significant 2.5 (1.4, 4.3). A case-only analysis revealed a near significant interaction between the two mutations among Caucasians. We found that clinical characteristics of VTE patients differed across race groups. Family history of VTE was common in white and black patients, yet known genetic risk factors for VTE are rare in African-American populations. Our findings underscore the need to determine gene polymorphisms associated with VTE in African-Americans.

Keywords: epidemiology, factor V Leiden, prothrombin, venous thromboembolism

Correspondence: Dr Nicole F. Dowling, Centers for Disease Control and Prevention, 1600 Clifton Road, N.E., Mailstop E-64, Atlanta, GA 30333, USA. Tel.: + 404 371 5261, fax: + 404 371 5424, e-mail: ndowling@cdc.gov

[1]Use of trade names is for identification only and does not constitute endorsement by the U.S. Department of Health and Human Services, the Public Health Service, or the Centers for Disease Control and Prevention.

Received 24 May 2002, accepted 5 August 2002

Venous thromboembolism (VTE) is a common vascular disease and significant public health problem in the USA, affecting about 1 in 1000 individuals per year [1]. Deep vein thrombosis (DVT), the most frequent presentation of VTE, is associated with significant morbidity and mortality. The most serious complication of DVT, pulmonary embolism (PE), is a life-threatening condition with short-term survival of less than 60% [2].

VTE is a multifactorial disease, resulting from a complex interaction of genetic and acquired factors. Primary hypercoagulability due to inherited deficiencies in anticoagulant proteins may be present in between 5 and 10% of patients with VTE [3]. Acquired factors such as surgery, malignancy, and immobilization have been associated with increased propensity for thrombosis [4,5]. While both acquired and inherited factors play important roles in the pathogenesis of VTE, risk varies greatly from one individual to another, and the causes for many cases remain unidentified. Moreover, little is known about the importance of interaction between environmental factors and inherited coagulation abnormalities in the development of VTE.

In the past decade, two single nucleotide polymorphisms, factor (F)V Leiden (G1691A) and prothrombin G20210A, have been demonstrated to be risk factors for DVT and PE in European and American Caucasians [6–11]. Despite these advances, the determinants of VTE in African-Americans are not well understood. Few cases of VTE in American blacks can be attributed to the FV Leiden or prothrombin G20210A mutations, because these variants are exceedingly rare among African-Americans. Prevalence of the FV allele, while at least 3% in American whites [11], is about 0.4% in healthy African-Americans [12]. Similarly, birth prevalence of the prothrombin 20210 A allele in African-Americans has been reported to be 0.2% [13], compared with an allele prevalence of 1% for healthy Caucasians [9]. Recent research indicates that incidence of idiopathic VTE in the USA may be higher for African-Americans than for Caucasians, yet inherited factors determining elevated VTE risk among blacks have not been elucidated [14]. Furthermore, the excess VTE risk for blacks has not been explained by elevated prevalence of other medical and surgical conditions.

To date, most epidemiologic studies of VTE risk factors have been conducted within white American and European populations. Moreover, few studies have evaluated interactive effects of these factors on risk of VTE. The goals of the present study

are to determine, in an American population, genetic and environmental factors associated with VTE, with a primary emphasis on elucidating race-specific aspects of the etiology and pathogenesis of this complex disease. In this paper, we present a detailed description of the study methods and results of an initial epidemiologic analysis of clinical and lifestyle characteristics of the study population. Additionally, we have investigated the associations of the FV Leiden, prothrombin G20210A, and 5,10-methylenetetrahydrofolate reductase (MTHFR) C677T variants with VTE by using several alternative statistical methods.

Materials and methods

Subject enrollment

The Genetic Attributes and Thrombosis Epidemiology (GATE) study is an ongoing case–control study of risk factors for VTE. Subject enrollment commenced in March 1998 and will continue though 2003. The study protocol was approved by the Institutional Review Boards at the participating institutions.

Patients, aged 18–70 years, hospitalized at two university-owned hospitals in Atlanta, Georgia with recently diagnosed first or recurrent episodes of DVT and/or PE are eligible as cases in the study. Potential cases are identified from a daily review of medical charts of all patients at the two hospitals receiving unfractionated or low-molecular-weight heparin. A DVT is objectively confirmed when diagnosed by Doppler ultrasonography, computed tomography (CT), magnetic resonance imaging, or contrast venography. Diagnosis of PE is made after positive angiogram, ventilation–perfusion lung scan, or CT. Patients with severe illness or with cognitive deficits, who are not able to complete study activities, are considered ineligible.

Control subjects were selected from a list of patients who visited the office of one of 10 physicians at a university-affiliated primary care clinic between January 1, 1997 and December 31, 2000. A patient list was obtained from the clinic's computerized patient accounts database. The master list was sampled to obtain a randomly ordered subset list of potential controls approximately similar to cases in age, sex, and race distributions. Individuals with a history of VTE, currently taking anticoagulant medication, or with a mental or physical problem precluding participation are not eligible to participate in the study.

Data collection

A whole blood sample is drawn from each hospitalized case subject and sent to a Centers for Disease Control and Prevention (CDC) laboratory for genetic analysis. Control subjects have a single blood sample, for genetic and plasma functional analysis, drawn at CDC at the time of enrollment. A second blood sample is obtained from case subjects, during a follow-up appointment at the CDC laboratory, at least 1 month after completion of anticoagulation therapy and at least 3 months after the index thrombotic event. The plasma samples will be used in future analyses to determine plasma levels of clotting factors, anticoagulant proteins, and other components affecting the coagulation system.

Each participant is interviewed at the time of enrollment. The 45-min interview takes place in the hospital for cases and at CDC for controls. Questions were designed to elucidate lifestyle, environmental, and medical factors that may be associated with VTE. Questions cover demographics, medical history, personal and family history of thrombosis, smoking and alcohol use, current medications, reproductive history and use of contraceptive and replacement hormones, physical activity, and diet and supplements. Specific questions are asked about life events and conditions occurring within the 4 weeks preceding VTE diagnosis for cases and within 4 weeks preceding enrollment in the study for controls. These life events and conditions include surgery, bed rest of more than 2 days duration, injury requiring medical treatment, travel of more than 8 hours' duration, and confinement to a wheelchair. Finally, a detailed medical records review is conducted for each enrolled case subject.

Enrollment numbers for current analysis

As of 1 March 2001, we identified 886 patients with DVT and/or PE at the two hospitals. Of these 886 patients, 147 were determined to be too ill to participate in the study and 20 others died before we could ask them about enrollment. Thirty-three additional patients were identified but had not yet been invited to participate. Of the 686 remaining patients, 387 (57%) agreed to enroll in the study, while 157 refused and 142 were lost to follow-up.

We sampled 616 control subjects from the patient lists of clinic physicians. Forty-two individuals were excluded because of history of VTE or current use of anticoagulant medications. Twelve were not contacted at the request of the physician. An additional 22 patients had not been contacted as of 1 March 2001. Of the 540 patients eligible to be control subjects in the study, 264 (49%) agreed to participate. A total of 151 patients refused and 125 individuals could not be located.

We excluded 12 cases and four controls with missing questionnaire or DNA data. As the present analysis includes only caucasian and African-American persons, five cases and 10 controls reporting a different racial background were excluded from the analytic file. The analysis for this paper includes 370 cases and 250 controls.

Laboratory methods

Blood samples were collected in 0.109 mol L^{-1} sodium citrate. DNA was extracted from the whole blood samples according to the manufacturer's protocol using the Puregene™ kit from Gentra Systems, Inc. (Minneapolis, MN, USA) and then stored at −20 °C. Allelic discrimination was used for DNA analysis as described by Benson and colleagues [15]. Polymerase chain reaction (PCR) primers and fluorogenic probes were designed and synthesized by PE/Applied Biosystems (Foster City, CA, USA) for the target regions. The probes were fluorescence-

labeled with reporter dyes of FAM (6-carboxyfluorescein, 6 FAM) and VIC™ on the 5′ ends for sequences determining mutation and wild-type, respectively. The probes were synthesized with a 3′-blocking phosphate, as well as a minor groove binder and a nonfluorescent quencher. PCR amplifications were performed in a GeneAmp PCR System 9600 (PE/Applied Biosystems). Final concentrations of reactant in a 20:1 mixture containing 10–100 ng of total DNA were 900 nmol L^{-1} of each primer, 200 mol L^{-1} of each probe, and 1X *Taq*Man® Universal Master Mix. Following an initial cycle of 50 °C for 2 min and 95 °C for 10 min, each cycle consisted of 92 °C for 15 s and 60 °C for 2 min for 40 amplification cycles. The *Taq*Man® assay was subsequently used for mutation detection.

Statistical methods

Goals of the statistical analyses are to evaluate the associations of gene polymorphisms and environmental factors with VTE and to assess gene-environment and gene–gene interactions. Odds ratios, 95% confidence intervals (CIs) and two-tailed *P*-values were obtained by large-sample methods (Mantel–Haenszel, unconditional logistic regression) computed by SAS version 8.1 software (SAS Institute, Cary, NC, USA) [16,17]. In cases where a cell expected value was ≤ 5, conditional maximum likelihood estimates for odds ratios and mid-P exact CIs and *P*-values were used [18]. Student's *t*-tests were used to test case–control differences between continuous variables.

For each gene polymorphism, genotypes were classified as homozygous wild-type or heterozygous or homozygous for the variant allele. In all comparisons, the homozygous wild-type genotype was considered the referent group. Odds ratios were calculated as the odds of being a case for each genotype divided by the odds of being a case for the referent genotype. The interpretation of the odds ratio is the relative risk of VTE for subjects with that genotype compared with subjects with the referent genotype. The χ^2 distribution was used to assess differences in allele frequencies between cases and controls and between racial groups. χ^2 tests were also used to test the assumption of Hardy–Weinberg equilibrium for each polymorphism.

We evaluated multiplicative, two-way interaction between homozygosity/heterozygosity for FV Leiden, homozygosity/heterozygosity for prothrombin G20210A, and homozygosity for the MTHFR T allele in a case-only analysis [19,20]. GATE data describing interaction of these genes were presented recently in a methods paper of the case-only design (L. Botto *et al.*, submitted for publication). Odds ratios bigger than unity in this analysis indicate more than a multiplicative effect of the two genetic factors. Case-only analysis requires that the exposures are statistically independent. Thus, we tested whether each of the three genetic factors was in linkage disequilibrium by obtaining a disequilibrium coefficient by maximum likelihood estimation [21]. Estimates in the vicinity of zero indicate no linkage disequilibrium and hence justify a critical assumption of the case-only analysis. We evaluated linkage disequilibrium among 4344 pooled control subjects obtained from the membership of a large health plan in California and enrolled at their annual physical examination [CDC controls, Thrombosis and Genes (TAG) study]. Our findings indicate the statistical independence of these genetic factors and support the use of case-only analyses.

We also evaluated two-way, multiplicative interaction between the three genetic factors using standard case–control methodology. However, these analyses were not possible using only GATE controls because the data are too sparse. Thus, we supplemented the GATE controls with the TAG study controls (assayed at the same laboratory as were GATE subjects). First, we evaluated whether or not the prevalence of the three genetic factors was statistically equivalent in the GATE and CDC TAG controls. We considered 'exposure' to be homozygosity or heterozygosity for the FV Leiden mutation, homozygosity or heterozygosity for the prothrombin G20210A variant, or homozygosity for the T allele of the MTHFR C677T polymorphism. The odds ratios of the 'exposed' genotypes for GATE controls compared with CDC TAG controls are 0.86, 1.6, and 0.75 for FV Leiden, prothrombin G20210A, and MTHFR C677T, respectively. No odds ratio is statistically significant (each *P*-value greater than 0.20). The similarity of these genetic factors of the GATE and CDC TAG controls provides a rationale for pooling these two groups.

Results

The average age of case subjects (49.2 years) is comparable to the average age of control subjects (49.5 years). A diagnosis of DVT was given to 250 case subjects, while 74 received a diagnosis of PE only, and 46 a diagnosis of PE with concomitant DVT. Of the cases, 255 (69%) were enrolled in the study when diagnosed with a first episode of VTE, while 115 (31%) reported at least one previously diagnosed VTE. Case–control comparisons first were conducted separately for subjects with a first episode of VTE and for those with recurrent VTE. As we found no statistically significant differences in association estimates between these two groups, we have reported results of analyses for all case subjects together.

As displayed in Table 1, cases are less likely than controls to be of caucasian race, to have a college degree ($P < 0.001$), and to have an annual household income greater than $40 000 ($P < 0.001$). Cases reported family history of VTE more frequently than control subjects (28% vs. 12%). Current smoking is not associated with risk of VTE, but alcohol consumption of up to seven drinks a week confers a statistically significant reduced risk of VTE. Body mass index (BMI) is higher for cases than for controls. This case–control difference, however, is seen only among Caucasians, with average body BMI for patients with VTE of 28.9 kg m^{-2} compared with 26.4 kg m^{-2} for control subjects ($P < 0.0001$). In logistic regression analysis, adjusting for age, gender, and education, a comparison of Caucasians with BMI of 30 kg m^{-2} (obese) to a referent group with normal BMI indicates a relative risk of 3.3 (95% CI 1.7, 6.1). In African-American subjects, case and control BMI averages are higher than corresponding values for Caucasians

Table 1 Characteristics of study participants

Characteristic	Cases (%) ($N=370$)	Controls (%) ($N=250$)	OR (95% CI)
Sex			
Male	181 (49)	125 (50)	1.0*
Female	189 (51)	125 (50)	1.0 (0.8, 1.4)*
Race			
African-American	174 (47)	91 (36)	1.0*
Caucasian	196 (53)	159 (64)	0.6 (0.5, 0.9)*
Annual income†			
<$24 999	134 (37)	29 (12)	1.0*
$25 000 – $39 999	79 (22)	47 (19)	0.4 (0.1, 0.9)*
$40 000 – $54 999	38 (10)	36 (14)	0.2 (0.1, 0.7)*
$55 000 – $70 000	39 (11)	39 (16)	0.2 (0.1, 0.6)*
>$70 000	73 (20)*	98 (39)	0.2 (0.1, 0.4)*
Education			
≤High school graduate	158 (43)	37 (15)	1.0*
Some college	96 (26)	60 (24)	0.4 (0.2, 0.6)*
Junior college degree	19 (5)	14 (6)	0.3 (0.1, 0.7)*
Four-year college degree	47 (13)	58 (23)	0.2 (0.1, 0.3)*
Postgraduate work	50 (13)	81 (32)	0.1 (0.09, 0.2)*
Smoking			
Not current	295 (80)	209 (84)	1.0‡
Current	75 (20)	41 (16)	1.0 (0.6, 1.5)‡
Alcohol use			
Rare/never	119 (32)	46 (18)	1.0‡
Light drinker	177 (48)	168 (67)	0.6 (0.4, 0.9)‡
Moderate drinker	43 (12)	22 (9)	0.9 (0.4, 1.7)‡
Heavy drinker	30 (8)	14 (6)	0.6 (0.3, 1.4)‡
Family history of VTE§			
No	220 (72)	208 (88)	1.0‡
Yes	87 (28)	29 (12)	2.3 (1.4, 3.8)‡
Body mass index ($kg\,m^{-2}$)			
<18.5	10 (3)	7 (3)	0.7 (0.2, 2.1)‡
18.5–24.9	91 (24)	81 (32)	1.0‡
25.0–30.0	126 (34)	89 (36)	1.3 (0.8, 2.0)‡
≥30.0	143 (39)	73 (29)	1.5 (0.9, 2.3)‡

*Odds ratios and 95% confidence intervals. †Income missing for one case subject. ‡Odds ratios, adjusted for age, gender, race, and education, and 95% confidence intervals. §Calculated excluding 76 subjects with 'unknown' family history.

but do not differ by case–control status ($29.6\,kg\,m^{-2}$ for cases vs. $30.6\,kg\,m^{2}$ for controls, $P>0.20$).

In Table 2, medical history factors are compared for cases and controls; results are calculated, adjusting for age, gender, race, and education. History of malignancy and of congestive heart failure are more frequent among subjects with VTE than among control subjects. History of diabetes and kidney failure are associated with elevated risk of VTE, but the results are not statistically significant. Case subjects are more likely than control subjects to report recent surgery, bed rest, injury, and confinement to a wheelchair.

We evaluated differences in medical history and clinical characteristics between caucasian and African-American case subjects (Table 3). While history of cancer is more prevalent among white cases (31% vs. 15%), kidney disease, hypertension, and diabetes (though not statistically significant) were reported more frequently by black VTE cases. Recent surgery and immobilization (bed rest and wheelchair use) are more common among white subjects with VTE. Family history of VTE was reported with equal frequency by white (29%) and black (28%) cases. The average age for black case subjects at time of index event diagnosis is 47.5 years, a value that is significantly lower than the average age of 50.7 years for white cases ($P=0.02$).

We calculated race-specific and race-adjusted odds ratios to assess the relationships between case–control status and FV Leiden, prothrombin G20210A, and MTHFR C677T genotypes (Table 4). Caucasians with the FV mutation have a three-fold increased risk of VTE; the relative risk for African-Americans is elevated but is not statistically significant. For whites, the prevalence of the prothrombin gene mutation is higher for cases than for controls (8.2% vs. 5.0%) but the difference does not reach statistical significance. Similarly, a race-adjusted odds ratio is above, though not statistically different from, unity; a relative risk among blacks could not be estimated because of the rarity of the mutation. The MTHFR variant is not associated with VTE among caucasian or African-American subjects. Frequencies of the FV Leiden A allele are 2.8% and 0.5% for whites and blacks, respectively. The prothrombin 20210 A allele is present in 2.5% of white control subjects but is absent in black controls. The T allele (MTHFR C6777T) prevalence also is highest in whites (35.5% vs. 11.5%).

The case-only analysis suggests a multiplicative, though not statistically significant, interaction between FV Leiden and prothrombin G20210A mutations on VTE risk (Table 5). There

Table 2 Medical history of study participants

	Cases (%) $N=370$	Controls (%) $N=250$	OR*	95% CI*
Chronic diseases				
Cancer	87 (24)	33 (13)	2.5	(1.6, 4.1)
Diabetes	64 (17)	23 (9)	1.5	(0.8, 2.5)
Hypertension	160 (43)	94 (38)	1.0	(0.7, 1.5)
Kidney disease	27 (7)	6 (2)	2.3	(0.9, 5.8)
Heart failure	31 (8)	2 (1)	9.2	(2.1, 39.8)
Conditions in 4 weeks preceding VTE diagnosis or preceding enrollment in study (controls)				
Surgery	140 (38)	3 (1)	50.8	(15.8, 163.4)
Bed rest > 2 days	211 (57)	10 (4)	29.4	(14.9, 58.1)
Injury	36 (10)	6 (2)	3.9	(1.6, 9.8)
Travel > 8 h	50 (14)	33 (13)	1.1	(0.7, 1.9)
Wheelchair	26 (7)	2 (1)	9.4	(2.2, 40.9)

*Odds ratios, adjusted for age, gender, race, and education, and 95% confidence intervals.

is evidence of interaction between prothrombin and MTHFR C677T variants on disease risk, but the effect is considerably smaller than that for FV Leiden and prothrombin. The joint effect of FV Leiden and MTHFR is smaller than the product of the marginal effects. The inclusion of the large number of CDC TAG study controls allows for direct evaluation of multiplicative interaction in a case–control analysis. This analysis supports the findings of the case-only analysis. The measure of multiplicative interaction between FV Leiden and the prothrombin mutation is somewhat increased in the case–control analysis, although the finding is not statistically significant ($P = 0.11$).

In addition, we evaluated the main effect of each genetic factor using the GATE cases compared with a combined group

Table 3 Comparison of characteristics of patients with VTE, by race

	Caucasians ($n = 196$)		African-Americans ($n = 174$)		
	Number	%	Number	%	*P*-value
Family history of VTE	48	29	39	28	>0.20
Diagnosis of PE	66	34	54	31	>0.20
First VTE	130	66	125	72	>0.20
Chronic diseases					
Cancer	60	31	27	15	0.001
Diabetes	28	14	36	21	0.10
Hypertension	75	38	85	49	0.04
Kidney disease	8	4	19	11	0.01
Heart failure	19	10	12	7	>0.20
Conditions in 4 weeks preceding VTE diagnosis					
Surgery	89	45	51	29	0.001
Bed rest > 2 days	126	64	85	49	0.003
Injury	22	11	14	8	>0.20
Travel > 8 h	32	16	18	10	0.09
Wheelchair	18	9	8	5	0.09
Mean age (standard deviation)	50.7 years	(12.7)	47.5 years	(12.7)	0.02

Table 4 Case–control comparison of genotype frequencies for factor V Leiden, prothrombin G20210A, and MTHFR C677T polymorphisms

	Caucasian				African-American			
	Cases	Controls			Cases	Controls		
Genotypes	$N = 194$	$N = 159$	OR*	95% CI*	$N = 174$	$N = 91$	OR*	95% CI*
Factor V Leiden								
G/G	85.5%	95.0%	1.0	–	97.1%	98.9%	1.0	–
G/A + A/A	14.5%	5.0%	3.2	(1.4, 7.2)	2.9%	1.1%	2.7	(0.4, 64.0)
Race-adjusted OR = 3.1 $(1.5, 6.7)_{95\%}$ *P*-value for homogeneity >0.20:								
Prothrombin G20210A								
G/G	91.8%	95.0%	1.0	–	98.9%	100%	1.0	–
G/A + A/A	8.2%	5.0%	1.7	(0.7, 4.1)	1.1%	0%	Inf.	–
Race-adjusted OR = 1.9 $(0.8, 4.4)_{95\%}$ *P*-value for homogeneity > 0.20:								
MTHFR C677T								
C/C	43.3%	37.7%	1.0	–	85.1%	76.9%	1.0	–
C/T	47.4%	53.5%	0.8	(0.5, 1.2)	13.2%	23.1%	0.5	(0.3, 1.0)
T/T	9.3%	8.8%	0.9	(0.4, 2.0)	1.7%	0%	Inf.	–

T/T vs. C/C + C/T: Race-adjusted OR = 1.2 $(0.6, 2.5)_{95\%}$ *P*-value for homogeneity > 0.20. *Odds ratios and 95% confidence intervals.

Table 5 Case-only and case–control analysis of two way multiplicative interaction between VTE, factor V Leiden, prothrombin mutation, and MTHFR polymorphisms among caucasian subjects

	Case-only analysis			Case–control analysis*		
	OR†	SE log(OR)‡	*P*-value	OR†	SE log(OR)‡	*P*-value
Interactions						
Factor V Leiden *prothrombin	2.37	0.6255	0.17	3.33	0.7531	0.11
Prothrombin*MTHFR	1.63	0.8094	>0.20	1.31	0.8417	>0.20
Factor V Leiden* MTHFR	0.69	0.7852	>0.20	0.67	0.8050	>0.20

*Pooled control subjects from GATE and TAG studies. †Odds ratio for multiplicative interaction. ‡Standard error of the logarithm of interaction odds ratio.

of GATE and CDC TAG study controls. The prothrombin G20210A mutation-VTE association is not statistically significant in the GATE-only analysis (Table 4). However, inclusion of the CDC TAG study controls increases the odds ratio to 2.5, and the finding is statistically significant ($P = 0.001$).

Discussion

The GATE study is the first large case–control study designed to evaluate genetic, environmental, and medical factors related to VTE in an American population. Our initial findings suggest that clinical characteristics of patients hospitalized with VTE differ significantly by race. Among caucasian participants, history of cancer, recent surgery, immobilization, injury, and heart failure are more prevalent among individuals with VTE, compared with control subjects. Heit and colleagues [4] noted that these factors are important risk factors for VTE within a predominantly caucasian population in Minnesota. For African-American VTE patients in our study, another set of characteristics is more common. Comparing black and white cases, we found higher frequencies of diabetes, hypertension, and kidney disease among blacks yet lower frequencies than among whites of conditions such as surgery, cancer, immobilization, and injury. For blacks, frequencies of diabetes and hypertension were nearly as high in controls as in cases, probably reflecting a higher rate of these chronic diseases in African-Americans overall. While we do not have sufficient evidence to conclude that these observed differences are related to the etiology of VTE, we can state that two different sets of clinical characteristics are present in white and black hospitalized VTE patients in the GATE study.

The results of our comparison, by race, of surgical history of VTE patients differ from the findings of another recent study of VTE. White and colleagues used California hospital discharge data from 1991 to 1994 to compare, across ethnic groups, incidence rates of VTE [14]. Within the California population, African-American patients with secondary VTE, compared with Caucasians with the same discharge diagnosis, had higher rates of colon and hip surgery preceding VTE diagnosis. In contrast, we noted the prevalence of all surgery to be significantly higher for white VTE patients (45%) than for black VTE patients (29%). However, we have not stratified our data by specific surgical procedures. Another source of differing results may be White's use of hospital discharge data, information subject to recording errors and misclassification. In the GATE study, diagnosis of VTE for case subjects was carefully ascertained by systematic review of hospital radiology reports.

Our study currently includes inpatients with VTE at two large hospitals. We did not enroll individuals diagnosed with VTE and treated as outpatients. Although not including outpatient VTE cases may limit the generalizability of our study findings, we do believe that the number of individuals diagnosed with VTE and treated outside of the hospital between 1998 and 2001 in our study area is quite small. Moreover, we have no reason to conclude that type of treatment for VTE differed substantially by race during this period. Another possible limitation of our study design is our use of outpatient clinic controls. Controls were, on average, more educated and had higher annual income than case subjects. We do not believe these differences will impact on the genetic findings of our study. For all other analyses of clinical and environmental factors, we have adjusted for education. Therefore, potential for bias is minimized.

We report associations for the FV Leiden, prothrombin G20210A, and MTHFR C677T polymorphisms that are similar to estimates reported in the recent literature. In the past decade, epidemiologic studies of caucasian populations have identified the FV Leiden mutation as a cause of primary [6,11,22] and recurrent VTE [23]. The A allele of the prothrombin G20210A polymorphism has been associated with elevated prothrombin levels and with a threefold increased risk of VTE in whites [9,10]. Studies of the C677T variant of the MTHFR gene, a polymorphism associated with blood levels of homocysteine [24], have yielded conflicting results [25,26]. In caucasian subjects in the GATE study, FV Leiden is strongly associated with VTE, while prothrombin 20210 A confers a nonsignificant elevation of VTE risk. We found both mutations to be rare and thus not important determinants of VTE among blacks. The MTHFR 677T allele is not related to VTE in participants of either race.

The case-only analysis suggests a multiplicative interaction between FV Leiden and the prothrombin G20210A variant on VTE risk. Typically, the evaluation of interaction between rare genetic traits in case–control studies is impossible because of lack of statistical power. In the present study, this interaction model failed using just GATE controls because none of these controls had both the FV Leiden and the prothrombin mutations. With respect to the evaluation of the interaction odds ratio, the case-only method substitutes the need for a control group with the assumption that the two exposures are independent. For genetic traits, this assumption can be evaluated easily by consideration of linkage disequilibrium, so that genetic studies are good candidates for case-only analyses. Through our use of the historical CDC controls, we were able to increase the number of controls considerably and could evaluate statistical interaction between the two genetic mutations using standard case–control methodology. This analysis provided support for the validity of our case-only analysis. However, we note that no matter how large the control group, the case-only analysis will provide a more precise estimate of the interaction odds ratio if the two genetic traits are independent. This fact is evidenced by the slightly larger standard errors of the interaction odds ratios for the case–control analysis compared with the case-only analysis in Table 5.

We believe that the use of historic (external) controls in genetic studies is under-utilized. In the present study, our pool of CDC controls enabled us to evaluate linkage disequilibrium with high statistical power. At the least, this evaluation provided a strong justification for the case-only study. Additionally, the observation that the GATE and CDC control groups did not differ significantly with respect to the distribution of these genetic factors provides justification for pooling control groups. The odds ratio for the prothrombin mutation based on the GATE

study becomes more positive and statistically significant using pooled controls. As we have no reason to suspect that the genetic background of whites in Atlanta is different from that of whites in California, and since our data suggest that the two groups do not differ on these genetic factors, we believe our study provides persuasive evidence that the prothrombin mutation is a cause of VTE.

Several recent studies have examined the effects of combinations of inherited risk factors on VTE risk among Caucasians. An interaction between hyperhomocysteinemia and the FV Leiden mutation has been reported by at least two research groups [27,28]. Cattaneo *et al.* [29] reported a statistically significant effect modification by the MTHFR C677T mutation on the association between the FV Leiden mutation and VTE. Findings from the Leiden Thrombophilia Study [25], however, did not support a role for the MTHFR variant in VTE risk among individuals with or without the FV Leiden mutation. Similarly, Brown *et al.* [30] and Alhenc-Gelas *et al.* [31] noted no significant interaction between the MTHFR C677T and either the FV Leiden or the prothrombin G20210A mutation. One recent study has demonstrated an increased risk for recurrent VTE among patients with both the FV Leiden and prothrombin G20210A mutations compared with the risk of recurrence among carriers of FV Leiden alone [32]. A second study analyzing pooled data from eight European case–control studies of VTE reported an odds ratio of 20.0 for double heterozygotes of FV Leiden and prothrombin mutations, a finding that represents a multiplicative interactive effect [33]. Our results, using data from a single case-series, suggest a multiplicative interaction between the FV Leiden and prothrombin G20210A mutations among Caucasians.

One of the more striking aspects of our analyses is that despite the rarity of known genetic risk factors among African-Americans, we found the prevalence of family history of VTE to be equal for black and white cases. This finding suggests that a strong genetic component exists in the etiology of VTE also among African-Americans. To date, the set of genetic factors responsible for a significant proportion of VTE cases among blacks remains undetermined. These results only underscore the need for research that addresses risk factors and etiologic mechanisms for VTE specific to an African-American population. Among both whites and blacks, an understanding of interactive effects between acquired and inherited factors is key to elucidating causes of VTE. The GATE study, with an ongoing enrollment of a large bi-racial study population, will provide the opportunity for a thorough evaluation of the complex etiology of VTE in Caucasians and African-Americans.

Acknowledgements

The authors thank Cathy Lally, Michele Beckman, Martha Forrester, Eileen Osinski, Pam Ruby, and Cathy Wood-Siverio (the GATE study staff); Penny Castellano, M.D and Kenneth Smith, M.D; and the CDC Hematologic Diseases Branch laboratory staff.

References

1 Silverstein MD, Heit JA, Mohr DN, Petterson TM, O'Fallon W. Melton LJ III. Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. *Arch Intern Med* 1998; **158**: 585–93.

2 Heit JA, Silverstein MD, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ III. Predictors of survival after deep vein thrombosis and pulmonary embolism: a population-based, cohort study. *Arch Intern Med* 1999; **159**: 445–53.

3 Heijboer H, Brandjes DP, Buller HR, Sturk A, ten Cate JW. Deficiencies of coagulation-inhibiting and fibrinolytic proteins in outpatients with deep-vein thrombosis. *N Engl J Med* 1990; **323**: 1512–6.

4 Heit JA, Silverstein MD, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ. Risk factors for deep vein thrombosis and pulmonary embolism. *Arch Intern Med* 2000; **160**: 809–15.

5 Goldhaber SZ. Epidemiology of pulmonary embolism and deep vein thrombosis. In. Bloom, AL Forbes, CD Thomas, DP Tuddenham, EGD, eds. *Haemostasis and Thrombosis, 3*. Edinburgh: Churchill Livingstone, 1994: 1327–33.

6 Koster T, Rosendaal FR, de Ronde H, Briet E, Vandenbroucke JP, Bertina RM. Venous thrombosis due to poor anticoagulant response to activated protein C: Leiden Thrombophilia Study. *Lancet* 1993; **342**: 1503–6.

7 Svensson PJ, Dahlback B. Resistance to activated protein C as a basis for venous thrombosis. *N Engl J Med* 1994; **330**: 517–22.

8 Ridker PM, Hennekens CH, Selhub J, Miletich JP, Malinow MR, Stampfer MJ. Interrelation of hyperhomocyst(e)inemia, factor V Leiden, and risk of future venous thromboembolism. *Circulation* 1997; **95**: 1777–82.

9 Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3′-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood* 1996; **88**: 3698–703.

10 Margaglione M, Brancaccio V, Giuliani N, D'Andrea G, Cappucci G, Iannaccone L *et al.* Increased risk for venous thrombosis in carriers of the prothrombin G→A20210 gene variant. *Ann Intern Med* 1998; **129**: 89–93.

11 Ridker PM, Hennekens CH, Lindpaintner K, Stampfer MJ, Eisenberg PR, Miletich JP. Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. *N Engl J Med* 1995; **332**: 912–7.

12 Hooper WC, Dilley A, Ribeiro MJ, Benson J, Austin H, Silva V *et al.* A racial difference in the prevalence of the Arg506→Gln mutation. *Thromb Res* 1996; **81**: 577–81.

13 Dilley A, Austin H, Hooper WC, El-Jamil M, Whitsett C, Wenger NK *et al.* Prevalence of the prothrombin 20210 G-to-A variant in blacks: infants, patients with venous thrombosis, patients with myocardial infarction, and control subjects. *J Lab Clin Med* 1998; **132**: 452–5.

14 White RH, Zhou H, Romano PS. Incidence of idiopathic deep venous thrombosis and secondary thromboembolism among ethnic groups in California. *Ann Intern Med* 1998; **128**: 737–40.

15 Benson JM, Ellingsen D, El-Jamil M, Jenkins M, Miller CH, Dilley A *et al.* Factor V Leiden and factor V, R2 allele: high–throughput analysis and association with venous thromboembolism. *Thromb Haemost* 2001; **86**: 1188–92.

16 Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; **22**: 719–48.

17 Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: John Wiley & Sons, Inc., 1989.

18 Martin D, Austin H. An efficient program for computing conditional maximum likelihood estimates and exact confidence limits for a common odds ratio. *Epidemiology* 1991; **2**: 359–62.

19 Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case–control studies. *Stat Med* 1994; **13**: 153–62.
20 Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene–environment interaction: case – control studies with no controls. *Am J Epidemiol* 1996; **144**: 207–13.
21 Weir B. *Genetic Data Analysis II*. Sunderland: MA. Sinauer Associates, Inc., 1996.
22 Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H *et al.* Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature* 1994; **369**: 64–7.
23 Simioni P, Prandoni P, Lensing AW, Scudeller A, Sardella C, Prins MH *et al.* The risk of recurrent venous thromboembolism in patients with an Arg506→Gln mutation in the gene for factor V (factor V Leiden). *N Engl J Med* 1997; **336**: 399–403.
24 Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG *et al.* A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* 1995; **10**: 111–3.
25 Kluijtmans LA, den Heijer M, Reitsma PH, Heil SG, Blom HJ, Rosendaal FR. Thermolabile methylenetetrahydrofolate reductase and factor V Leiden in the risk of deep-vein thrombosis. *Thromb Haemost* 1998; **79**: 254–8.
26 Margaglione M, D'Andrea G, d'Addedda M, Giuliani N, Cappucci G, Iannaccone L *et al.* The methylenetetrahydrofolate reductase TT677 genotype is associated with venous thrombosis independently of the coexistence of the FV Leiden and the prothrombin A20210 mutation. *Thromb Haemost* 1998; **79**: 907–11.
27 Mandel H, Brenner B, Berant M, Rosenberg N, Lanir N, Jakobs C *et al.* Coexistence of hereditary homocyteinuria and factor V Leiden – effect on thrombosis. *N Engl J Med* 1996; **334**: 763–8.
28 Ridker PM, Hennekens CH, Selhub J, Miletich JP, Malinow MR, Stampfer MJ. Interrelation of hyperhomocyst(e) inemia, factor V Leiden, and the risk of future venous thromboemoblism. *Circulation* 1997; **95**: 1777–82.
29 Cattaneo M, Tsai MY, Bucciarelli P, Taioli E, Zighetti M, Bignell M *et al.* A common mutation in the methylenetetrahydrofolate reductase gene (C677T) increases the risk for deep-vein thrombosis in patients with mutant factor V (Factor V. Q^{506}). *Arterioscler Thromb Vasc Biol* 1997; **17**: 1662–6.
30 Brown K, Luddington R, Baglin T. Effect of the MTHFRC677T variant on risk of venous thromboembolism: interaction with factor V Leiden and prothrombin (F2G20210A) mutations. *Br J Haematol* 1998; **103**: 42–4.
31 Alhenc-Gelas M, Arnaud E, Nicaud V, Aubry M, Fiessinger J, Aiach M *et al.* Venous thromboembolic disease and the prothrombin, methylene tetrahydrofolate reductase and factor V genes. *Thromb Haemost* 1999; **81**: 506–10.
32 De Stefano V, Martinelli I, Mannucci PM, Paciaroni K, Chiusolo P, Casorelli I *et al.* The risk of recurrent deep venous thrombosis among heterozygous carriers of both factor V Leiden and the G20210A prothrombin mutation. *N Engl J Med* 1999; **341**: 801–6.
33 Emmerich J, Rosendaal FR, Cattaneo M, Margaglione M, DeStefano V, Cumming T *et al.* Combined effect of factor V Leiden and prothrombin 20210A on the risk of venous thromboembolism. *Thromb Haemost* 2001; **86**: 809–16.

Genetic Variants of Arachidonate 5-Lipoxygenase–Activating Protein, and Risk of Incident Myocardial Infarction and Ischemic Stroke

A Nested Case-Control Approach

Robert Y.L. Zee, PhD; Suzanne Cheng, PhD; Hillary H Hegener, BS; Henry A. Erlich, PhD; Paul M Ridker, MD

Background and Purpose—Recent findings have implicated specific gene polymorphisms of arachidonate 5-lipoxygenase–activating protein (ALOX5AP), and 2 at-risk haplotypes (HapA, HapB) in myocardial infarction and stroke. To date, no prospective data are available.

Methods—We evaluated 10 specific Icelandic ALOX5AP gene variants among 600 male participants with incident atherothrombotic events (myocardial infarction [MI] or ischemic stroke) and among 600 age- and smoking-matched male participants, all white, who remained free of reported cardiovascular disease during follow-up within the Physicians' Health Study cohort.

Results—Overall allele, genotype, and haplotype distributions were similar between cases and controls. Single-marker conditional logistic regression analysis adjusted for potential risk factors found no association with risk of atherothrombotic events. Further investigation using a haplotype-based approach showed similar null findings with MI (HapA: odds ratio [OR]=1.18, 95% CI, 0.76 to 1.85; P=0.46; HapB: odds ratio=0.62, 95% CI, 0.36 to 1.07; P=0.08), and with ischemic stroke (HapA: odds ratio=1.11, 95% CI, 0.65 to 1.89; P=0.71; HapB: odds ratio=0.82, 95% CI, 0.47 to 1.42; P=0.47).

Conclusions—We found no evidence for an association of the specific Icelandic ALOX5P gene variants/at-risk haplotypes tested with risk of incident MI nor ischemic stroke in this prospective, non-Icelandic study. (***Stroke***. **2006;37:2007-2011.**)

Key Words: ALOX5AP ■ haplotypes ■ MI ■ risk factors ■ stroke

Cardiovascular diseases, including myocardial infarction (MI) and ischemic stroke, are the leading causes of mortality and morbidity in western countries. The underlying pathogenesis is likely to be mediated by both genetic and environmental risk factors. The initial report,[1] in an Icelandic population, of a significant association of genetic variants of arachidonate 5-lipoxygenase–activating protein (ALOX5AP) with increased risk of MI and stroke has attracted great interest. In their study, Helgadottir and coauthors reported a linkage and association of a 4–single-nucleotide polymorphism (SNP) haplotype, HapA, of *ALOX5AP* gene with risk of MI and stroke.[1] In addition, they reported an association of a different 4-SNP haplotype, HapB, with risk of MI in a British population.[1] Helgadottir and coauthors further assessed the contribution of *ALOX5AP* variants, in particular the HapA, and HapB haplotypes, to stroke, in a Scottish population, and found that the HapA haplotype confers a relative risk of 1.36 assuming a multiplicative model (P=0.007) for stroke.[2] However, they found no association for HapB. Subsequent studies by others in several non-Icelandic populations have since yielded conflicting results.[3,4]

To date, no prospective genetic-epidemiological data are available on risk of MI, and ischemic stroke. We therefore simultaneously evaluated the role of 10 ALOX5AP (GeneID: 241; Chromosome: 13q12) SNPs (SG13S25, SG13S377, SG13S106, SG13S114, SG13S89, SG13S30, SG13S32, SG13S41, SG13S42, and SG13S35), and specific haplotypes thereof, in particular HapA, and HapB at-risk haplotypes, as risk determinants of incident MI, and ischemic stroke in a prospective, nested case-control sample within the Physicians' Health Study (PHS) cohort. These polymorphisms (except SG13S106, SG13S30, and SG13S42: unpublished data from deCODE Genetics) were chosen based on the associations observed in the Icelandic study.[1]

Materials and Methods

Study Design

We used a nested case-control design within the PHS,[5] a randomized, double-blinded, placebo-controlled trial of aspirin and beta carotene initiated in 1982 among 22 071 males, predominantly white

Received March 31, 2006; accepted May 2, 2006.

From the Center for Cardiovascular Disease Prevention, the Donald W. Reynolds Center for Cardiovascular Research, the Leducq Center for Molecular and Genetic Epidemiology (R.Y.L.Z., H.H.H., P.M.R.), Brigham and Women's Hospital, Harvard Medical School, Boston, Mass; and the Roche Molecular Systems (S.C., H.A.E.), Alameda, Calif.

Correspondence to Dr Robert Y.L. Zee, Laboratory of Genetic and Molecular Epidemiology, Center for Cardiovascular Disease Prevention, Brigham and Women's Hospital, 900 Commonwealth Ave East, Boston, MA 02215. E-mail rzee@rics.bwh.harvard.edu

Stroke is available at http://www.strokeaha.org

DOI: 10.1161/01.STR.0000229905.25080.01

EXHIBIT E

(>94%), US physicians, 40 to 84 years of age at study entry. Before randomization, 14 916 participants provided an EDTA-anticoagulated blood sample and stored for genetic analysis. All participants were free of prior MI, stroke, transient ischemic attacks, and cancer at study entry. As the study participants were all US male physicians, yearly follow-up self-report questionnaires provide reliable updated information on newly developed diseases and the presence or absence of other cardiovascular risk factors. History of cardiovascular risk factors, such as hypertension (>140/90 mm Hg or on antihypertensive medication), diabetes or hyperlipidemia (>240 mg/dL), was defined by self-report of diagnosis at entry into the study. For all reported incident vascular events occurring after study enrollment, hospital records, death certificates, and autopsy reports were requested and reviewed by an end-points committee using standardized diagnostic criteria.

The diagnosis of MI was confirmed by evidence of symptoms in the presence of either diagnostic elevations of cardiac enzymes or diagnostic changes on electrocardiograms. In the case of fatal events, the diagnosis of MI was also accepted based on autopsy findings. Stroke was defined by the presence of a new focal neurological deficit, with symptoms and signs persisting for >24 hours, and was ascertained from blinded review of medical records, autopsy results and the judgment of a board-certified neurologist, on the basis of clinical reports, computed tomographic, or MRI scanning.

For each case (MI or ischemic stroke), a control matched by age, smoking history (never, past, or current) and length of follow-up were chosen among those subjects who remained free of vascular diseases. The present association study consisted of 341 MI case-control pairs, and 259 ischemic stroke case-control pairs, all white males.

The study was approved by the Brigham and Women's Hospital Institutional Review Board for Human Subjects Research.

Genotyping Determination

Genotyping was performed using an immobilized probe approach, as previously described (Roche Molecular Systems).[6] In brief, each DNA sample was amplified in a multiplex polymerase chain reaction using biotinylated primers. Each polymerase chain reaction product pool was then hybridized to a panel of sequence-specific oligonucleotide probes immobilized in a linear array. The colorimetric detection method was based on the use of streptavidin-horseradish peroxidase conjugate with hydrogen peroxide and 3,3',5,5'-tetramethylbenzidine as substrates.

To confirm genotype assignment, scoring was carried out by 2 independent observers. Discordant results (<1% of all scoring) were resolved by a joint reading, and where necessary, a repeat genotyping. Results were scored blinded as to case-control status. Overall completion rate of genotyping determination was ≥95%.

Statistical Analysis

Allele and genotype frequencies among cases and controls were compared with values predicted by Hardy-Weinberg equilibrium using the χ^2 test. Relative risks associated with each genotype were calculated separately by conditional logistic regression analysis conditioning on the matching by age, smoking status, and length of follow-up since randomization, and further controlling for randomized treatment assignment, history of hypertension, presence or absence of diabetes, and body mass index, assuming an additive, dominant, or recessive mode of inheritance. Pairwise linkage disequilibrium (LD) was examined as described by Devlin and Risch.[7] For comparison with published reports by others, we examined 2 previously described at-risk haplotypes: HapA (SG13S25*G*-SG13S114*T*-SG13S89*G*-SG13S32*A*), and HapB (SG13S377*A*-SG13S114*A*-SG13S41*A*-SG13S35*G*). Haplotype estimation and inference was determined using PHASE v2.1.[8,9] Haplotype distributions between cases and controls were examined by likelihood ratio test. The relationship between haplotypes and clinical outcomes was examined using a haplotype-based logistic regression analysis with baseline-parameterization,[10] adjusting for the same risk factors. All analyses were carried out using SAS/Genetics 9.1 package (SAS Institute, Inc). For each odds ratio (OR), we calculated 95% CIs. A 2-tailed P value of 0.05 was considered a statistically significant result.

TABLE 1. Baseline Characteristics of Study Participants Who Subsequently Developed Any Arterial Event (Cases), and Those Who Remained Free of Vascular Disease During Follow-Up (Controls)

	Controls (n=600)	Cases (n=600)	P
Age, y	60.8±0.3	61.0±0.3	m.v.
Smoking status, %			m.v.
Never	41.7	41.7	
Past	41.5	41.5	
Current	16.8	16.8	
Body mass index, kg/m²	24.9±0.1	25.4±0.1	0.002
Blood pressure, mm Hg			
Systolic	128.6±0.5	132.7±0.6	<0.0001
Diastolic	79.6±0.3	81.8±0.3	<0.0001
Hyperlipidemia, %	14.9	22.8	<0.001
Hypertension, %	29.0	47.2	<0.0001
Diabetes, %	2.8	8.9	<0.0001
Aspirin use, %	46.3	44.8	0.61
Family history of premature CAD <60 years of age, %	8.9	10.9	0.24

Mean±SE unless otherwise stated.

m.v. indicates matching variable; CAD, coronary artery disease.

Continuous and categorical variables were tested by paired t test and McNemar test, respectively.

Results

Baseline characteristics of cases and controls are shown in Table 1. As expected, the case participants had a higher prevalence of traditional cardiovascular risk factors at baseline as compared with controls. The genotype frequencies for the polymorphisms tested were in Hardy-Weinberg equilibrium in the control group and in the case group.

Using a single-marker χ^2 analysis, allele and genotype distributions were similar between cases and controls (Table 2). Results from the adjusted conditional logistic regression analysis, assuming additive, dominant, or recessive mode of inheritance, showed no significant association of the variants tested with the clinical outcomes (P≥0.07; data not shown). In general, the polymorphisms tested were in LD (supplemental Table I, available online at http://stroke.ahajournals.org). The overall haplotype distributions between cases and controls were similar (MI: HapA region, P=0.79, HapB region, P=0.94; ischemic stroke: HapA region, P=0.77, HapB region, P=0.26; supplemental Table II, available online at http://stroke.ahajournals.org). The most frequent haplotypes were G-T-G-C, and G-T-A-G for HapA region, and HapB region, respectively (supplemental Table II), and thus were used as the referents. Results from the adjusted haplotype-based conditional logistic regression analysis again showed similar null findings (supplemental Table III, available online at http://stroke.ahajournals.org).

TABLE 2. Genotype and Allele Distribution

ALOX5AP Genotype, %	MI Controls	MI Cases	*P*	IsST Controls	IsST Cases	*P*
SG13S25			0.80			0.29
GG	81.31	80.56		83.13	79.58	
GA	18.07	19.14		15.64	20.00	
AA	0.62	0.31		1.23	0.42	
Allele			0.89			0.47
G	0.90	0.90		0.91	0.90	
A	0.10	0.10		0.09	0.10	
SG13S377			0.71			0.35
GG	75.39	78.09		70.37	75.42	
GA	23.05	20.68		25.93	22.50	
AA	1.56	1.23		3.70	2.08	
Allele			0.41			0.15
G	0.87	0.88		0.83	0.87	
A	0.13	0.12		0.17	0.13	
SG13S106			0.54			0.20
GG	50.16	46.60		45.27	45.00	
GA	37.69	41.98		44.86	40.00	
AA	12.15	11.42		9.88	15.00	
Allele			0.59			0.38
G	0.69	0.68		0.68	0.65	
A	0.31	0.32		0.32	0.35	
SG13S114			0.90			0.96
TT	47.04	45.37		41.56	42.08	
TA	41.43	42.28		43.62	42.50	
AA	11.53	12.35		14.81	15.42	
Allele			0.63			0.99
T	0.68	0.68		0.63	0.63	
A	0.32	0.32		0.37	0.37	
SG13S89			0.76			0.80
GG	89.72	88.89		89.71	89.17	
GA	9.66	10.80		9.47	10.42	
AA	0.62	0.31		0.82	0.42	
Allele			0.84			0.96
G	0.95	0.94		0.94	0.94	
A	0.05	0.06		0.06	0.06	
SG13S30			0.83			0.38
GG	58.57	58.95		51.85	57.92	
GT	37.69	36.42		41.15	36.67	
TT	3.74	4.63		7.00	5.42	
Allele			0.91			0.17
G	0.77	0.77		0.72	0.76	
T	0.23	0.23		0.28	0.24	
SG13S32			0.30			0.32
CC	27.73	22.84		24.28	20.83	
CA	52.96	54.63		47.33	54.17	
AA	19.31	22.53		28.40	25.00	
Allele			0.15			0.99
C	0.54	0.50		0.48	0.48	
A	0.46	0.50		0.52	0.52	

(Continued)

TABLE 2. Continued

ALOX5AP Genotype, %	MI Controls	MI Cases	*P*	IsST Controls	IsST Cases	*P*
SG13S41			0.50			0.89
AA	82.87	83.02		84.36	85.42	
AG	15.58	16.36		14.40	13.75	
GG	1.56	0.62		1.23	0.83	
Allele			0.73			0.68
A	0.91	0.91		0.92	0.92	
G	0.09	0.09		0.08	0.08	
SG13S42			0.17			0.36
AA	28.04	34.88		38.68	35.00	
AG	50.78	45.99		43.62	50.00	
GG	21.18	19.14		17.70	15.00	
Allele			0.11			0.88
A	0.53	0.58		0.60	0.60	
G	0.47	0.42		0.40	0.40	
SG13S35			0.08			0.50
GG	81.31	85.80		79.42	83.33	
GA	18.69	13.58		19.75	16.25	
AA	...	0.62		0.82	0.42	
Allele			0.21			0.26
G	0.91	0.93		0.89	0.91	
A	0.09	0.07		0.11	0.09	

IsST indicates ischemic stroke.
P value for χ^2 test.

Discussion

The present prospective investigation provides no evidence for an association of the specific gene variants, nor at-risk haplotypes of the *ALOX5AP* gene, previously suggested as genetic risk determinants, with MI or stroke in a non-Icelandic white population.

In the initial Icelandic report,[1] a 4-SNP haplotype (HapA) was found to be associated with a 2× greater risk of MI, and an almost 2× greater risk of stroke. The same group also reported an association of a different 4-SNP ALOX5AP haplotype (HapB) with risk of MI in a British sample population[1] (Table 3). A subsequent report by Helgadottir and coauthors found an association between HapA and an increased risk of ischemic stroke (relative risk=1.35; *P*=0.02), and an over-representation of HapB (relative risk=1.65; *P*=0.02) with ischemic stroke in a Scottish male sample population[2] (Table 3). Recently, Lohmussar and coauthors[3] reported that sequence variants in the *ALOX5AP* gene are significantly associated with stroke, particularly in males, in a Central European sample population. A nominally significant association with stroke was observed for SG13S114 (OR=1.24; *P*=0.017), and SG13S100 (OR=1.26; *P*=0.024). However, they found no association of HapA with stroke risk.[3] More recently, Meschia and coauthors conducted the first replication study using a North American sample

TABLE 3. Summary of ALOX5AP At-Risk-Haplotypes Association Studies

	HapA		HapB	
	MI	Stroke	MI	Stroke
	Conf, Casf, R, *P*	Conf, Casf, R, *P*	Conf, Casf, R, *P*	Conf, Casf, R, *P*
Present study United States	0.14, 0.17, 1.18, 0.46	0.18, 0.15, 1.11, 0.71	0.07, 0.06, 0.62, 0.08	0.08, 0.07, 0.82, 0.47
Iceland[1]	0.10, 0.16, 1.80, <0.0001	0.10, 0.15, 1.67, <0.0001	Not available	*0.07, 0.07, 1.09, ns
United Kingdom[1]	0.15, 0.17, ns	Not available	0.04, 0.08, 1.95, 0.00037	Not available
Scotland[2]	Not available	0.14, 0.18, 1.35, 0.02	Not available	0.06, 0.09, 1.65, 0.02
Germany[3]	Not available	0.15, 0.15, ns	Not available	ns (data not shown)
North America[4]	Not available	ns (data not shown)	Not available	Not available

Conf indicates haplotype frequency in controls; Casf, haplotype frequency in cases; R, risk estimate; ns, nonsignificance.
HapA=SG13S25*G*-SG13S114*T*-SG13S89*G*-SG13S32*A*. HapB=SG13S377*A*-SG13S114*A*-SG13S41*A*-SG13S35*G*.
*Data extracted from reference 2.

population, and found no association between ALOX5AP gene variants and stroke, although MI was not investigated in their study.

Given this situation, a possible explanation for the apparent discrepancies is that the observed allele, genotype, and at-risk haplotype frequencies for the SNPs examined may differ between studies, which could be the result of population/ethnic differences. As previously suggested,[3,4] the *ALOX5AP* gene variation may play a substantial role in risk of MI, and stroke in Iceland (an isolate population), but a lesser role in non-Icelandic populations because of different population LD structures. These recent results are consistent with the initial report that different at-risk haplotypes were found between the Icelandic and British study populations.[1]

As shown in Table 3, not all of the published reports examined the same set of SNPs, nor did all of the reported studies examine the association of *ALOX5AP* variants with MI and stroke simultaneously. Further, not all published studies presented information on allele, genotype and at-risk haplotype frequencies, LD structure, and risk estimates, thus making a direct comparison and informative interpretation across studies difficult.

It has been noted in the initial report[1] that variants of *ALOX5AP* gene are involved in the pathophysiology of MI and stroke by increasing the production of leukotriene B4, a critical regulator in the 5-lipoxygenase pathway, and a proinflammatory agent. Leukotrienes are arachidonic acid metabolites, which have been implicated in various inflammatory conditions, including asthma, arthritis, psoriasis, and atherosclerosis.[11,12] Notably, a recent article by the same Icelandic group found a haplotype (HapK) of the gene encoding leukotriene A4 hydrolase, a protein in the same biochemical pathway of ALOX5AP, confers ethnicity-specific (particularly in blacks) risk of MI.[13]

The prospective nature of the PHS study and the use of a closed population sampling scheme in which subsequent case status was determined solely by the development of disease strongly reduce the possibility that our findings are attributable to bias or confounding. Our study cohort consists of entirely white males with distinct socioeconomic status (physicians), so our data cannot be generalized to other ethnic groups and women. In our study, we had the ability to detect, based on the present sample sizes, assuming 80% power, at an α of 0.05, a risk ratio of >1.54 (MI), and 1.64 (ischemic stroke) if the minor allele frequency is 0.50, and of >2.26 (MI), and 2.49 (ischemic stroke) if the minor allele frequency is 0.05 assuming a univariable-additive mode. Thus, we cannot rule out a modest risk of cardiovascular disease associated with the polymorphisms/haplotypes tested. It is important to recognize that association studies like this one can only examine the possible association between phenotype and the tested polymorphisms. Our study therefore cannot exclude the possibility that examination of different polymorphisms/loci, which would by definition have to be in linkage disequilibrium with the ones tested, might obtain different results.

In conclusion, our prospective study found no evidence for an association of specific Icelandic *ALOX5AP* gene polymorphisms/at-risk haplotypes examined with risk of atherothrombotic events. If corroborated in other non-Icelandic prospective studies, our data suggest that *ALOX5AP* gene variation is not informative for risk assessment of atherothrombosis in non-Icelandic populations.

Acknowledgments

We thank Michael Grow and Houman Khakpour (both at Roche Molecular Systems) for their expertise and efforts in developing the genotyping reagents used for this study. We thank Anna Helgadottir at deCODE Genetics for sharing unpublished data.

Sources of Funding

This work was supported by grants from the National Heart Lung and Blood Institute (HL-58755 and HL-63293), the Doris Duke Charitable Foundation, the American Heart Association, and the Donald W. Reynolds Foundation, Las Vegas, Nevada.

Disclosures

None.

References

1. Helgadottir A, Manolescu A, Thorleifsson G, Gretarsdottir S, Jonsdottir H, Thorsteinsdottir U, Samani NJ, Gudmundsson G, Grant SF, Thorgeirsson G, Sveinbjornsdottir S, Valdimarsson EM, Matthiasson SE, Johannsson H, Gudmundsdottir O, Gurney ME, Sainz J, Thorhallsdottir M, Andresdottir M, Frigge ML, Topol EJ, Kong A, Gudnason V, Hakonarson H, Gulcher JR, Stefansson K. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat Genet*. 2004;36:233–239.
2. Helgadottir A, Gretarsdottir S, St Clair D, Manolescu A, Cheung J, Thorleifsson G, Pasdar A, Grant SF, Whalley LJ, Hakonarson H, Thorsteinsdottir U, Kong A, Gulcher J, Stefansson K, MacLeod MJ. Association between the gene encoding 5-lipoxygenase-activating protein and stroke replicated in a scottish population. *Am J Hum Genet*. 2005;76:505–509.
3. Lohmussaar E, Gschwendtner A, Mueller JC, Org T, Wichmann E, Hamann G, Meitinger T, Dichgans M. *AlOX5AP* gene and the *PDE4D* gene in a central European population of stroke patients. *Stroke*. 2005;36:731–736.
4. Meschia JF, Brott TG, Brown RD Jr, Crook R, Worrall BB, Kissela B, Brown WM, Rich SS, Case LD, Evans EW, Hague S, Singleton A, Hardy J. Phosphodiesterase 4d and 5-lipoxygenase activating protein in ischemic stroke. *Ann Neurol*. 2005;58:351–361.
5. Physician's health study. Aspirin and primary prevention of coronary heart disease. *N Engl J Med*. 1989;321:1825–1828.
6. Cheng S, Grow MA, Pallaud C, Klitz W, Erlich HA, Visvikis S, Chen JJ, Pullinger CR, Malloy MJ, Siest G, Kane JP. A multilocus genotyping assay for candidate markers of cardiovascular disease risk. *Genome Res*. 1999;9:936–949.
7. Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics*. 1995;29:311–322.
8. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 2001;68:978–989.
9. Stephens M, Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet*. 2003;73:1162–1169.
10. Wallenstein S, Hodge SE, Weston A. Logistic regression model for analyzing extended haplotype data. *Genet Epidemiol*. 1998;15:173–181.
11. Samuelsson B. Leukotrienes: mediators of immediate hypersensitivity reactions and inflammation. *Science*. 1983;220:568–575.
12. Spanbroek R, Grabner R, Lotzer K, Hildner M, Urbach A, Ruhling K, Moos MP, Kaiser B, Cohnert TU, Wahlers T, Zieske A, Plenz G, Robenek H, Salbach P, Kuhn H, Radmark O, Samuelsson B, Habenicht AJ. Expanding expression of the 5-lipoxygenase pathway within the arterial wall during human atherogenesis. *Proc Natl Acad Sci U S A*. 2003;100:1238–1243.
13. Helgadottir A, Manolescu A, Helgason A, Thorleifsson G, Thorsteinsdottir U, Gudbjartsson DF, Gretarsdottir S, Magnusson KP, Gudmundsson G, Hicks A, Jonsson T, Grant SF, Sainz J, O'Brien SJ, Sveinbjornsdottir S, Valdimarsson EM, Matthiasson SE, Levey AI, Abramson JL, Reilly MP, Vaccarino V, Wolfe ML, Gudnason V, Quyyumi AA, Topol EJ, Rader DJ, Thorgeirsson G, Gulcher JR, Hakonarson H, Kong A, Stefansson K. A variant of the gene encoding leukotriene a4 hydrolase confers ethnicity-specific risk of myocardial infarction. *Nat Genet*. 2006;38:68–74.

****
****

Document-ID: CI-06652876-5

Patron: 30847/2048-004

Note:

NOTICE:

****
****

Pages: 11 Printed: 01-19-07 08:13:41

Sender: Ariel/Windows

nature genetics

The gene encoding phosphodiesterase 4D confers risk of ischemic stroke

Solveig Gretarsdottir[1], Gudmar Thorleifsson[1], Sigridur Th Reynisdottir[1], Andrei Manolescu[1], Sif Jonsdottir[1], Thorbjörg Jonsdottir[1], Thorunn Gudmundsdottir[1], Sigrun M Bjarnadottir[1], Olafur B Einarsson[1], Herdis M Gudjonsdottir[1], Malcolm Hawkins[1], Gudmundur Gudmundsson[1], Hrefna Gudmundsdottir[1], Hjalti Andrason[1], Asta S Gudmundsdottir[1], Matthildur Sigurdardottir[1], Thomas T Chou[1], Joseph Nahmias[1], Shyamali Goss[1], Sigurlaug Sveinbjörnsdottir[2], Einar M Valdimarsson[2], Finnbogi Jakobsson[2], Uggi Agnarsson[2], Vilmundur Gudnason[3], Gudmundur Thorgeirsson[3], Jurgen Fingerle[4], Mark Gurney[1], Daniel Gudbjartsson[1], Michael L Frigge[1], Augustine Kong[1], Kari Stefansson[1,5] & Jeffrey R Gulcher[1,5]

We previously mapped susceptibility to stroke to chromosome 5q12. Here we finely mapped this locus and tested it for association with stroke. We found the strongest association in the gene encoding phosphodiesterase 4D (*PDE4D*), especially for carotid and cardiogenic stroke, the forms of stroke related to atherosclerosis. Notably, we found that haplotypes can be classified into three distinct groups: wild-type, at-risk and protective. We also observed a substantial disregulation of multiple *PDE4D* isoforms in affected individuals. We propose that *PDE4D* is involved in the pathogenesis of stroke, possibly through atherosclerosis, which is the primary pathological process underlying ischemic stroke.

Stroke is a common and serious disease; each year in the United States more than 600,000 individuals suffer a stroke and more than 160,000 die from stroke-related causes[1]. In western countries, stroke is the leading cause of severe disability and the third leading cause of death[2]. The clinical phenotype of stroke is complex but is broadly divided into ischemic (accounting for 80–90%) and hemorrhagic stroke (10–20%; ref. 3). Ischemic stroke is further subdivided into large vessel occlusive disease (herein referred to as carotid stroke) commonly due to atherosclerotic involvement of the common and internal carotid arteries; small vessel occlusive disease, thought to be a nonatherosclerotic narrowing of small end-arteries in the brain; and cardiogenic stroke due to blood clots arising from the heart typically on the background of atrial fibrillation or ischemic (atherosclerotic) heart disease[4,5]. Therefore, stroke does not seem to be one disease but rather a heterogeneous group of disorders reflecting differences in the pathogenic mechanisms[6,7]. All forms of stroke share risk factors, such as hypertension, diabetes, hyperlipidemia and smoking[1,8]. Family history of stroke is also an independent risk factor, suggesting the existence of genetic factors that may interact with environmental factors[7,9].

The genetic determinants of the common forms of stroke are still largely unknown. There are examples of mutations in specific genes that cause rare mendelian forms of stroke[10–16], but none of these occur on the background of atherosclerosis, and, therefore, the corresponding genes are probably not involved in the common forms of stroke, which most often occur with atherosclerosis.

The first main locus associated with stroke, *STRK1*, was mapped to 5q12 using a genome-wide search for susceptibility genes in the common forms of stroke[17]. A broad but rigorous definition of the phenotype was used, including individuals that had ischemic stroke, transient ischemic attack (TIA) and hemorrhagic stroke. The lod score after adding a high density of markers (one marker per centimorgan) was 4.40 (P value = 3.9×10^{-6}) at marker *D5S2080*.

We describe here the positional cloning of a gene associated with susceptibility to stroke in the *STRK1* locus. We finely mapped the region and tested it for association to stroke, and we found the strongest association in *PDE4D*, encoding phosphodiesterase 4D, a member of the large superfamily of cyclic nucleotide phosphodiesterases. *PDE4D* was most strongly associated with the combination of two forms of stroke related to atherosclerosis: cardiogenic and carotid stroke. Relative expression of *PDE4D* isoforms correlated with stroke and correlated with the genetic variation of stroke associated with *PDE4D*.

RESULTS

Microsatellite allelic association

We initially genotyped 864 Icelandic affected individuals and 908 controls using a total of 98 microsatellite markers. These markers are

[1]deCODE Genetics, Sturlugata 8, IS-101 Reykjavik, Iceland. [2]National University Hospital, Reykjavik, Iceland. [3]Icelandic Heart Association Heart Preventive Clinic, Reykjavik, Iceland. [4]Hoffmann-La Roche, Pharma Division, Preclinical Research, Basel, Switzerland. [5]These authors contributed equally to this work. Correspondence should be addressed to S. Gretarsdottir (solveig.gretarsdottir@decode.is) or J.R.G. (jeffrey.gulcher@decode.is).

Published online 21 September 2003; doi:10.1038/ng1245

EXHIBIT F

distributed over a region of approximately 11 Mb. The region is centered on our linkage peak and corresponds to the 2-lod drop. The density of markers is greater in the central 3.7-Mb portion of the region, which includes the 1-lod drop, with an average spacing of one marker every 53 kb. We have designated this central region, which is flanked by markers *D5S1474* and *D5S398*, the *STRK1* interval. Three markers, AC027322-5, *D5S2121* and AC008818-1 had different allelic frequencies in affected individuals versus controls with values of $P < 0.01$ (**Table 1**). Correcting for the relatedness of the affected individuals had little impact on the *P* values, but after correcting for the number of markers and alleles tested, none of these *P* values were significant (**Table 1**).

We had previously observed that our linkage peak increased, though not significantly, when we excluded those affected with hemorrhagic stroke. We therefore also tested those affected with ischemic stroke or TIA for association to the markers. In addition, those affected with ischemic stroke and TIA were subclassified according to the TOAST research criteria, and we repeated the association analysis separately for the three TOAST subcategories: cardiogenic, carotid and small vessel occlusive disease. Finally, we tested the combination of those affected with either cardiogenic or carotid stroke, as these categories of stroke are most clearly related to atherosclerosis. The results for each of these association studies are presented in **Supplementary Table 1** online. Three of the markers were significantly associated, one for cardiogenic stroke (AC008818-1), one for carotid stroke (*DG5S397*) and one for the combination of carotid and cardiogenic stroke (AC008818-1), even after correcting for multiple testing (**Table 1**). The marker *DG5S397* is located within *PDE4D*; AC008818-1 is in the 5′ end of *PDE4D* and in the overlapping gene *PART1* (prostate androgen-regulated transcript), whose transcript is on the other strand going in the opposite direction. **Supplementary Fig. 1** online shows the locations of these and other markers relative to the genes in the *STRK1* interval.

PDE4D is an important regulator of intracellular levels of cAMP and is expressed widely. *PART1* encodes a putative protein with unknown function highly expressed in prostate and several tumor cell lines. Physical locations of all genotyped markers and *PDE4D* and *PART1* exons are available in **Supplementary Table 2** online. The association results for the combination of carotid and cardiogenic stroke were particularly notable, with an allele frequency for allele 0 (the CEPH reference allele) of marker AC008818-1 of 35.5% in affected individuals versus 25.5% in controls. The unadjusted *P* value for this marker is 0.0000015; after adjusting for multiple testing of markers, the *P* value is 0.00025 (**Table 1**). This is significant even after adjusting for the several phenotypes studied. The risk of this allele to the other alleles of this marker, assuming the multiplicative model[18,19], was estimated to be 1.60 with a corresponding population attributable risk of 25%. Thus, the strong association signals from our initial microsatellite association studies helped to focus our attention on the *STRK1* interval and, in particular, on the *PDE4D* gene region.

Screening for polymorphisms in *PDE4D*

We next considered whether a functional variant in *PDE4D* might be the cause of our observed microsatellite association. We matched

Table 1 Microsatellite and SNP allelic association

Phenotype	Marker	Allele	*P* value	*P* value[a]	*P* value[b]	RR	# Aff.	% Aff.	# Control	% Control
All affected	AC027322-5	10	0.0010	0.0012	NS	3.34	787	1.9	779	0.6
	D5S2121	-2	0.0027	0.0034	NS	2.19	824	2.7	870	1.3
	AC008818-1	0	0.0045	0.0050	NS	1.25	815	29.9	891	25.5
Cardiogenic	AC008818-1	0	0.000054	0.000077	0.011	1.60	216	35.4	891	25.5
	D5S1990	20	0.00053	0.00088	NS	2.18	223	7.9	879	3.8
	D5S2089	-10	0.0027	0.0040	NS	2.22	219	5.9	813	2.8
Carotid	*DG5S397*	4	0.00024	0.00031	0.045	1.70	124	65.7	577	53.0
	DG5S2056	12	0.00091	0.0019	NS	3.33	80	8.8	464	2.8
	AC008818-1	0	0.0010	0.0014	NS	1.61	125	35.6	891	25.5
Combined cardiogenic	AC008818-1	0	0.0000015	0.0000024	0.00025	1.60	341	35.5	891	25.5
and carotid	AC008833-6	0	0.0026	0.0032	NS	1.35	335	70.3	868	63.8
	DG5S2066	0	0.0032	0.0039	NS	1.74	258	92.3	501	87.2
All affected	SNP32	C	0.00024	0.00027	NS	1.46	400	37.9	475	29.5
	SNP56	T	0.0028	0.0031	NS	1.31	550	71.4	615	65.5
	SNP45	G	0.0065	0.0077	NS	1.33	723	82.4	492	78.0
Cardiogenic	SNP89	A	0.00023	0.00031	NS	2.10	150	90.0	450	81.1
	SNP45	G	0.00041	0.00053	NS	1.77	196	86.2	492	77.9
	SNP91	G	0.00047	0.00059	NS	2.02	151	89.7	451	81.3
Carotid	SNP83	C	0.00043	0.00053	0.045	1.94	76	67.8	349	52.0
	SNP87	T	0.00058	0.00063	NS	1.74	96	62.0	583	48.4
	SNP100	T	0.0010	0.0012	NS	1.79	99	36.4	339	24.2
Combined cardiogenic	SNP45	G	0.000034	0.000044	0.005	1.77	309	86.3	492	78.0
and carotid	SNP41	A	0.000078	0.000096	0.011	1.86	236	86.0	368	76.8
	SNP87	T	0.00019	0.00026	0.031	1.49	263	58.2	583	48.4
	SNP89	A	0.00025	0.00030	0.037	1.85	232	88.8	450	81.1
	SNP56	T	0.00027	0.00034	0.041	1.56	230	74.8	615	65.5

[a]*P* values adjusted for the relatedness of the affected. [b]*P* values adjusted for all the markers tested.

Presented in the table are the three most significant single-marker association results for the disease categories and all signals that survive correction for multiple testing. This is shown both for microsatellites (upper part) and for SNPs (lower part). For the microsatellites, the number reported as an allele is the offset from the smaller of the two alleles of CEPH sample 1347-02 (CEPH genomic repository); thus, allele 0 serves as a (CEPH) reference allele.

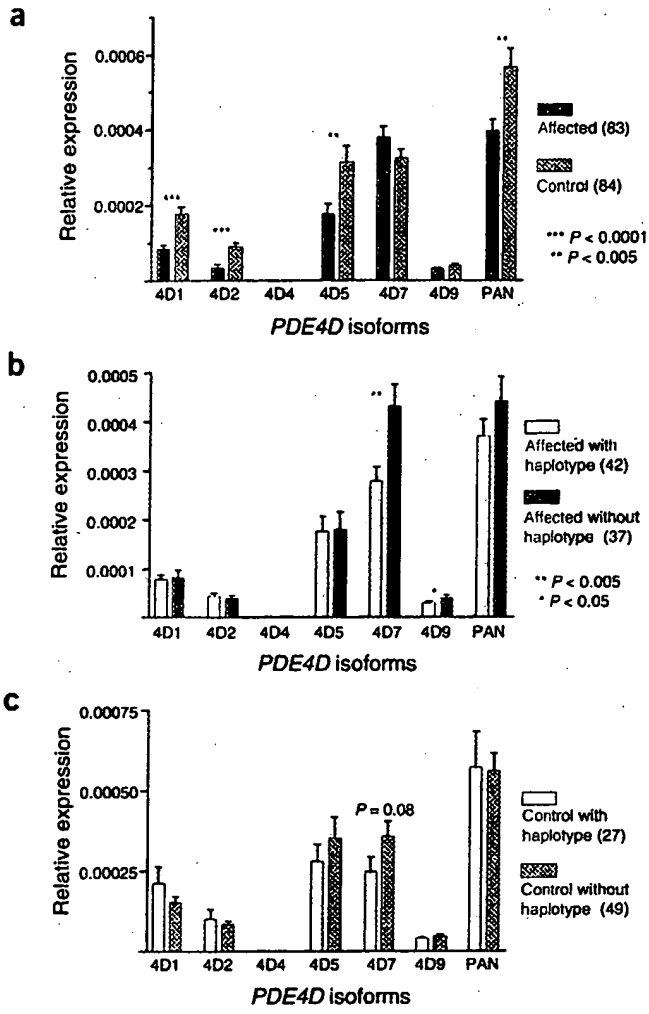


Figure 1 Expression of *PDE4D* isoforms in affected individuals and controls. Expression of *PDE4D* is shown relative to the expression of *GAPD* (as a housekeeping gene). The difference in expression between cases and controls was tested using a two-sample *t*-test on the log-transformed values. Two-sided *P* values are reported. Number of samples is given in parentheses. PAN, total expression of all isoforms. (**a**) Isoform-specific expression of *PDE4D* mRNA from a randomly selected cohort of affected individuals (red) and controls (blue). (**b**,**c**) Corresponding analysis comparing affected individuals (**b**) or controls (**c**) with (white bars) and without (colored bars) the at-risk haplotype G0 at the 5′ end of the gene.

public domain expressed-sequence tags and our own RT–PCR and RACE transcripts to our sequence of the *STRK1* interval and defined new alternative *PDE4D* transcripts (**Supplementary Note** online). *PDE4D* contains at least 22 exons over approximately 1.5 Mb overlapping with *PART1*. It encodes eight protein isoforms and has at least seven promoters. All isoforms identified have an identical C-terminal catalytic domain but differ at the N-terminal regulatory domain (**Supplementary Fig. 2** online).

We then attempted to identify mutations by sequencing all known *PDE4D* exons (including the overlapping *PART1* exons) and, on average, 100 bp of their flanking introns in 188 individuals affected with stroke and 94 controls. We identified 46 polymorphisms: 44 single-nucleotide polymorphisms (SNPs) and two intronic deletions. Only two of the polymorphisms, both SNPs, were found within the coding exons of *PDE4D*, consistent with the extraordinary lack of variation that others have reported for all four PDE4 classes[20]. We genotyped the two coding SNPs in additional affected individuals and controls, but they did not show significant association to stroke (**Supplementary Table 3** online). Therefore, if a functional variant conferring risk for stroke exists in *PDE4D*, it may be located in regulatory regions affecting transcription, splicing, message stability or message transport of one or more isoforms or in exons that we have not yet identified.

PDE4D isoform expression

Because we found no functional mutations in the known coding exons of *PDE4D*, we considered other evidence for this gene underlying the association in this region. We studied the expression levels of the various *PDE4D* isoforms, as significant differences between affected individuals and controls could indicate that regulation of *PDE4D* is a key element in stroke susceptibility. We used EBV-transformed B-cell lines from randomly selected affected individuals with ischemic stroke or TIA and from controls. We carried out isoform-specific kinetic RT–PCR analysis to quantify each isoform in 83 individuals with stroke and 84 controls. Most of the affected individuals had ischemic stroke, and 38% had cardiogenic or carotid stroke. We observed that the total *PDE4D* message level, as assessed by amplification across exons present in all isoforms, was significantly lower in affected individuals than in controls ($P = 0.0021$). This difference was due primarily to lower expression of the *PDE4D1*, *PDE4D2* and *PDE4D5* isoforms (**Fig. 1a**). This significant disregulation of the expression of multiple *PDE4D* isoforms encouraged us to continue investigating the association of *PDE4D* with stroke.

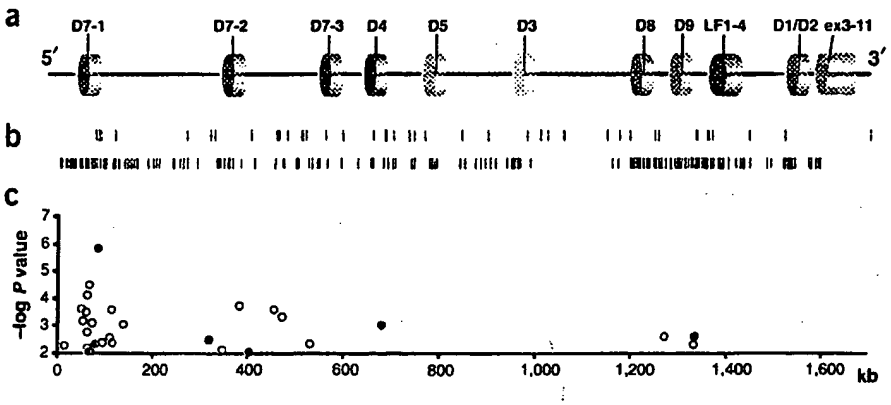


Figure 2 Single-marker allelic association within *PDE4D*. The same horizontal scales are used for **a**, **b** and **c**. (**a**) *PDE4D* gene structure. Exons are shown as colored cylinders and exon names are indicated above the line. (**b**) Microsatellite and SNP distribution in the gene. Red vertical bars indicate microsatellites and blue vertical bars SNPs. (**c**) Single-marker allelic association across *PDE4D* for both microsatellites (filled circles) and SNPs (open circles). The plot shows negative log *P* value versus the physical location in kilobases. Results with *P* values of 0.01 or less are shown for all stroke cases (black) and for the combination of cardiogenic and carotid cases (red).

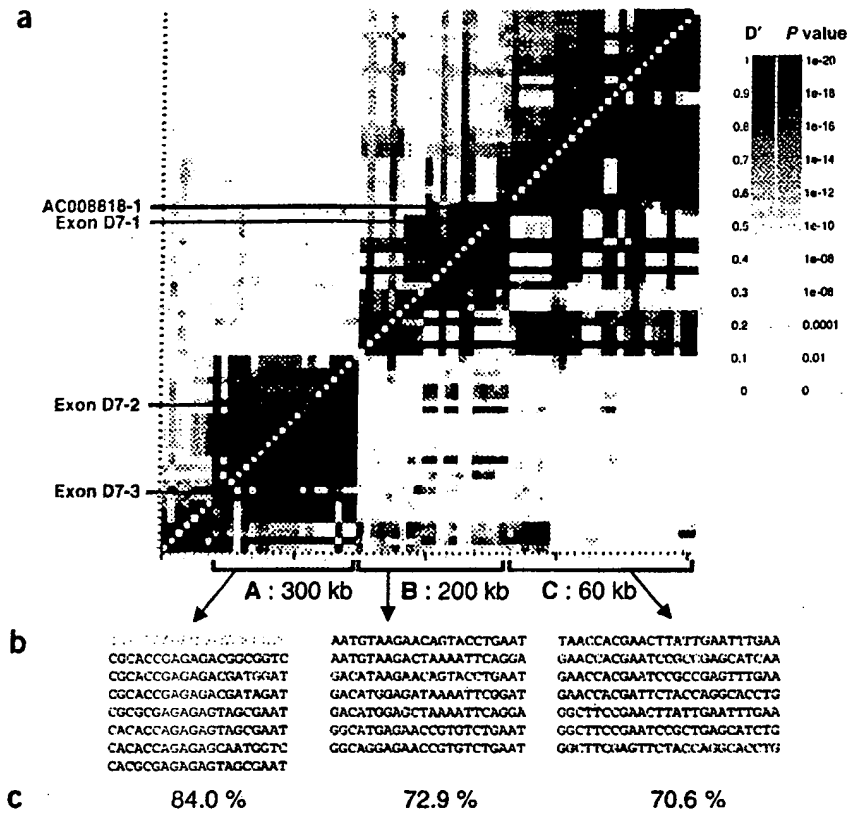


Figure 3 LD and haplotypes at the 5′ end of *PDE4D*. **(a)** Pairwise linkage disequilibrium between SNPs in a 600-kb region in the 5′ end of *PDE4D*. The markers are plotted equidistantly. Two measures of LD are shown: *D′* in the upper left triangle and *P* values in the lower right triangle. This region can be divided into three blocks of strong LD, each with limited haplotype diversity: block A, block B and block C. Colored lines indicate the position of the three exons, D7-1, D7-2 and D7-3, and the microsatellite marker, AC008818-1. **(b)** All common haplotypes identified in each of the three blocks. The haplotypes in each block showing strongest association with stroke are colored green, blue and red. Association results for all haplotypes are presented in **Supplementary Table 6** online. **(c)** The percentage of chromosomes in each block that match one of the common haplotypes.

SNPs: marker association and linkage disequilibrium

We next searched for SNPs in the intronic and flanking regions of *PDE4D* in the public National Center for Biotechnology Information SNP database or by sequencing selected intronic and flanking regions in the gene in at least 94 affected individuals and 94 controls. We initially identified 637 SNPS. Many of these SNPs were completely correlated so we removed several redundant SNPs from further genotyping. Some SNPs with very low minor allele frequencies were also ignored. This resulted in a set of 260 SNPs that were then genotyped in the entire affected and control cohorts. We determined the exonic structure of *PDE4D* (**Fig. 2a**) relative to the location of SNPs and microsatellite markers (**Fig. 2b**) and carried out single-marker SNP and microsatellite association tests for all markers (**Fig. 2c**).

Most markers with significant associations were located at the 5′ end of the gene. One SNP (SNP83) associated with carotid stroke and five of the SNPs (SNP45, SNP41, SNP87, SNP89 and SNP56) associated with the combined cardiogenic and carotid stroke were significant even after adjusting for all the SNPs tested (**Table 1**). Three of these significant SNPs flank exon D7-1; the other three are in a 100-kb region containing exon D7-2 (for physical positions see **Supplementary Table 2** online). Some additional results for the single-point SNP associations are supplied in **Supplementary Table 4** online.

The two most significant SNPs, SNP45 and SNP41, are within 6 kb of the microsatellite marker AC008818-1, and the at-risk alleles of all three genetic markers are in strong linkage disequilibrium (LD) with $D' > 0.9$ and *P* value nearly zero (**Supplementary Table 5** online). The square of the correlation (R^2) is very high between the two SNPs (~0.93) but is substantially lower (~0.08) between each SNP and the at-risk allele of the microsatellite. This is because the frequency of the at-risk alleles of the two SNPs are similar and much higher than that of the at-risk allele of the microsatellite. We determined the LD block structure around the 5′ end of *PDE4D* (**Fig. 3a**). We delineated three blocks, A, B and C, encompassing the first three exons of *PDE4D* and its immediate upstream region. Exons D7-3 and D7-2 are both in block A, and D7-1 (the first exon) is in block B, close to its border with block C. Given this block structure, we were prepared to investigate haplotypes associated with susceptibility to stroke in this region.

Haplotype association

We first considered haplotypes based on the most significantly associated SNPs and microsatellite, SNP45, SNP41 and AC008818-1, all in block B separated by only 6 kb. As expected given the high degree of correlation between SNP45 and SNP41, we found that it was sufficient to consider only the two marker haplotypes consisting of the microsatellite and SNP45, the SNP with the higher genotype yield. The results of this association study for the combination of carotid and cardiogenic stroke are shown in **Figure 4a**. The letter X designates the joint set of alleles, excluding the at-risk allele 0, of microsatellite AC008818-1. GX is therefore the composite of all haplotypes including the G nucleotide of SNP45 except for the G0 haplotype. For our samples, the A0 haplotype does not exist. This suggests that allele 0 originated in a haplotype background with allele G of SNP45 and since then, no recombination has occurred between those two markers for chromosomes that carried allele 0.

Haplotypes AX, G0 and GX carry significantly different risks for the combined carotid and cardiogenic stroke phenotype. We consider haplotype GX to be the wild type as it is the most common (53.4% in controls) and also because it carries an intermediate level risk not too different from the population risk. Haplotype G0 carries higher risk and haplotype AX is protective, with risks of 1.46 and 0.70 relative to the wild type, respectively. The risk associated with haplotype G0 is 2.07 times that of the protective haplotype AX. Each of the three pairwise comparisons was highly significant, with *P* values ranging from 0.006 to 7.2×10^{-8}. Both haplotypes AX and GX are composite haplotypes, but the AX haplotype can be simply summarized by the allele A of SNP45, as the haplotype A0 does not exist. Similarly, the G0 haplotype is completely determined by the 0 allele of AC008818-1.

Figure 4a also shows the information content (Info) of each test. The difference between Info and 1 is a measure of the information that is lost owing to the uncertainty with phase and missing geno-

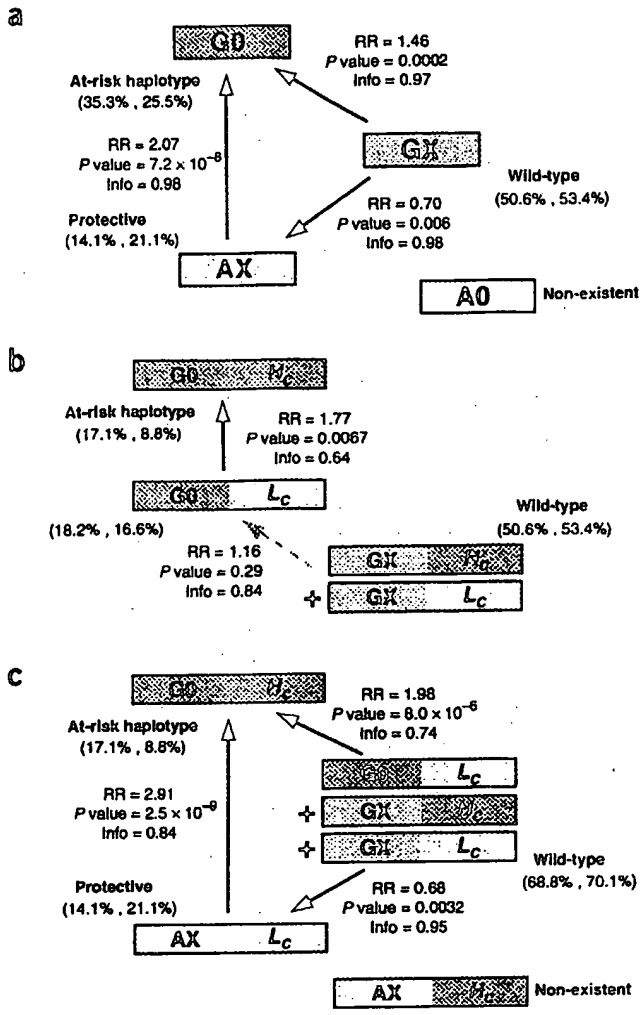


Figure 4 Haplotype association for carotid and cardiogenic stroke combined. Estimated haplotype frequencies for affected individuals and controls, in that order, are given in parentheses. (**a**) Comparisons of groups of haplotypes constructed from SNP45 and AC008818-1, two markers separated by 6 kb. X is a composite allele denoting all alleles of AC008818-1 except allele 0. Apart from haplotype A0, which is not found in our samples, other haplotypes can be grouped into three groups with distinct risks. Each arrow corresponds to a comparison between two groups, and RR is the estimated risk of the group the arrow is pointing at relative to the other group. (**b**) Intermediate results when the investigation is extended from SNP45 and AC008818-1, which are both in LD block B, to include 25 SNPs in LD block C. H_C is the at-risk haplotype, identified in **Figure 3** (colored in red), and L_C is a composite haplotype denoting all haplotypes of the 25 SNPs except H_C. Together with AC008818-1 and SNP45, the haplotypes here span 64 kb. Haplotype G0 in **a** is split into extended haplotypes G0H_C and G0L_C. G0H_C has significantly higher risk than G0L_C, and the risk of G0L_C is not distinguishable from that of the wild-type GX. (**c**) A refinement of the groupings in **a**. G0L_C is moved from the at-risk group to the wild-type group. The extended haplotype AXH_C does not exist, indicating that blocks B and C are in LD.

types (see **Supplementary Note** online for details). Info is very close to 1 for each of the three pairwise comparisons (**Fig. 4a**). This is a result of SNP45 and AC008818-1 being in very strong LD. Tests presented in **Figure 4b,c**, which involve longer haplotypes, have lower information content.

We next identified and estimated the risks for the common SNP haplotypes in each block, considering only those SNPs with minor allele frequency greater than 20%. Block A (300 kb) contained 19 such SNPs, block B (200 kb) 22 SNPs and block C (60 kb) 25 SNPs. We identified all haplotypes in each block with an estimated frequency in the population of 2% or greater. In each block there were fewer than ten such haplotypes, and they accounted for approximately 80% of the total haplotype frequency for that block. A brief schematic of the identified haplotypes is given in **Figure 3b**, and the risks and frequencies of these haplotypes are available in **Supplementary Table 6** online. In block A, no common haplotype has greater risk than SNP87 alone. The strongest signals were for haplotypes in block B and C. Each block contained a haplotype significantly associated with the combination of carotid and cardiogenic stroke and having relative risk around 1.5. The common at-risk haplotype in block B is the SNP background of the G0 haplotype previously identified.

Although there were no significant single-marker associations in block C, we observed a common haplotype with 15.4% frequency in controls, which we designate haplotype H_C. All haplotypes defined by the 25 SNPs in block C that are not H_C are jointly denoted by the composite haplotype L_C. We investigated the contribution of H_C in conjunction with the SNP45 and AC008818-1 haplotypes. AX and H_C do not exist together on the same chromosome (**Fig. 4c**), at least in these samples. Thus, blocks B and C are far from being independent, and the extended composite haplotype AXL_C is the same as AX. The haplotype G0 can be split into the two extended haplotypes G0H_C and the composite G0L_C, which have significantly different risks ($P = 0.0067$; **Fig. 4b**). Moreover, the high risk associated with G0 is totally accounted for by G0H_C, as G0L_C has risk that is not significantly different from GX (GX = GXH_C + GXL_C; **Fig. 4b**). This observation allowed us to refine our initial haplotype groupings (**Fig. 3a,c**). The extended at-risk haplotype G0H_C (8.8% in controls) and protective composite haplotype AXL_C (21.1% in controls) have relative risks of 1.98 and 0.68, respectively, relative to the wild type (70.1% in controls). Based on these risk estimates, if everybody's risk corresponded to that of a homozygous carrier of the protective variant, the number of cases would be reduced by 55%, which can be interpreted as the population-attributed risk of the at-risk haplotype and the wild-type combined.

The at-risk haplotype G0H_C spans a region of about 64 kb. It is possible that the greater risk is due to multiple polymorphisms over that region, but the results are also consistent with this region harboring a relatively recent mutation (as yet unidentified) that occurred in that haplotype background, with no recombination occurring since then for chromosomes carrying the mutation. By contrast, the protective composite haplotype AXL_C can be simply represented by allele A of SNP45. Hence, it is possible that allele A of SNP45 is the functional protective variant, although it is possible that the functional variant is simply in strong LD with allele A of SNP45 and has not yet been identified. Statistically, the effects of SNP45 and SNP41 are indistinguishable from each other.

We reanalyzed the *PDE4D* isoform expression data for those with haplotype G0 versus those without that haplotype both in affected individuals and in controls. For the samples in the expression study, the frequency of the G0 haplotype was 29.4% in affected individuals and 25.2% in controls. Those affected with the haplotype had significantly lower expression of the *PDE4D7* and *PDE4D9* isoforms (**Fig. 1b**). Other isoforms of *PED4D* did not significantly correlate with the disease-associated haplotype. The correlation of *PDE4D7* with the haplotype was also present in controls but was only marginally significant (**Fig. 1c**).

DISCUSSION

Our results indicate that variations in *PDE4D* are associated with ischemic stroke. The direct involvement of *PDE4D* is strongly supported by linkage in conjunction with association and expression analysis. We first identified the association using microsatellite markers and then supplemented the microsatellite data with a denser set of SNPs. The strongest association was with the two ischemic subtypes, carotid and cardiogenic stroke. We examined whether the disease-associated alleles and haplotype were related to specific stroke risk factors, such as hypertension, hypercholesterolemia, diabetes, peripheral artery occlusive disease and coronary artery disease in addition to early onset of stroke and sex (**Supplementary Table 7** online). We observed a marginally significant association to hypercholesterolemia, but the contribution of *PDE4D* to stroke is clearly not strongly correlated with any of these known risk factors.

For the combined cardiogenic and carotid subtype of stroke, it is notable that haplotypes covering the first exon of *PDE4D* can be classified into three groups with clearly distinct risks. Relative to the protective group, the general population-attributed risk of the at-risk and wild-type groups combined is estimated to be 55%. Approximately 16% of the general population carries one copy of the at-risk haplotype (Fig. 4c). They have about 1.8 times higher risk than the general population for cardiogenic or carotid stroke. Approximately 0.8% of the population are homozygous with respect to the at-risk haplotype and, assuming the multiplicative model, their risk is estimated to be about 3.8 times than that of the general population. We have not yet identified the functional variants that are responsible for the observed effects of these haplotype groups. And, because these haplotype groups do not fully explain the linkage signal we observe in the region for all affected individuals, we certainly could not rule out, and indeed expect, that there are other variants or haplotypes in *PDE4D* not directly related to those we have identified that confer risk to stroke. These are probably rare but could have very high penetrance. We also cannot rule out the possibility that some other genes in the linkage region independent of, or in conjunction with, *PDE4D* confer susceptibility to stroke.

By alternative splicing and using different promoters, *PDE4D* generates at least eight different isoforms that yield functional proteins, differing from each other at their N-terminal regions. We identified four new exons encoding the N-termini of two new isoforms, *PDE4D7* and *PDE4D9*. The disease-associated haplotype extends over the 5′ exon unique to the new *PDE4D7* variant and the presumed promoter region of this isoform, suggesting that the functional variation may be involved in transcriptional regulation. This hypothesis is also supported by our *PDE4D* expression analysis showing that there is significant correlation between the disease-associated haplotype and the level of *PDE4D7* message.

The strongest association found for this *PDE4D* haplotype was to the two main subtypes of ischemic stroke, cardiogenic and carotid stroke, suggesting a role for this gene in the vascular biology of atherosclerosis. Although there are multiple etiologies for ischemic stroke, atherosclerosis is the most important and is the primary pathological process for cardiogenic and carotid stroke. First, it is the main cause of stenotic and occlusive lesions of the internal and common carotids that lead to carotid strokes. Second, cardiac thrombi, which shed emboli to the brain, most commonly occur on the background of coronary artery disease (such as after acute myocardial infarction or ischemic cardiomyopathy) or as a result of atrial fibrillation due to poor compliance of ischemic ventricles (diastolic dysfunction/stiffening). Although atrial fibrillation may occur on the background of other diseases, such as valvular disease, hyperthyroidism and hypertension, in the age group that tends to suffer from stroke, ischemic heart disease is one of the main causes. Ischemic stroke resulting from occlusion of small penetrating arteries in the brain (small vessel occlusive disease) is generally thought to result from endothelial proliferation, as atherosclerosis only occurs in larger arteries. *PDE4D* does not show association to small vessel stroke, consistent with its role in atherosclerosis.

What biological role does *PDE4D* have in predisposition to stroke, in particular, and to the underlying atherosclerosis? PDE4D selectively degrades second messenger cAMP[20], which has a central role in signal transduction and regulation of physiological responses. It is expressed in most cell types important to the pathogenesis of atherosclerosis, including vascular smooth muscle cells, endothelial cells, T-lymphocytes, macrophages[21–25] and monocytes (data not shown). Cyclic AMP is a key signaling molecule in these cells[26–28]. In vascular smooth muscle cells, low cAMP levels lead to an increase in proliferation and migration that is mediated, at least in part, by PDE4 (refs. 26,29,30). Animal models have also shown that elevation of cAMP reduces neointimal lesion formation and inhibits proliferation of smooth muscle cells after arterial injury[31,32]. In monocytes and T-lymphocytes, accumulation of cAMP is generally associated with inhibition of immune functions, such as proliferation and cytokine secretion[33].

One could postulate that the regulation of cAMP through absolute or relative expression of one or more *PDE4D* isoforms may differ in individuals susceptible to stroke; some may have greater PDE4D activity and, consequently, lower cAMP levels in any of the above cell types, leading to development of the atherosclerotic plaque or to its instability. Contrary to what one might expect, however, we observed lower expression of some of the *PDE4D* isoforms in EBV cell lines from affected individuals. These isoforms are upregulated by cAMP[22,34,35], suggesting disregulation at the level of cAMP in affected individuals. It is therefore possible that greater activity of one or few splice variants alters the effective PDE4D enzymatic activity of the cell, decreasing the cAMP levels and thus altering the expression of cAMP-regulated isoforms as observed in our expression study. This relative expression of *PDE4D* isoforms may determine the compartmental localization of *PDE4D* isoforms and thus the corresponding gradients of intracellular cAMP that have been recently observed[20].

In summary, we present association analyses (single-marker and haplotype analyses) that support the idea that *PDE4D* confers risk of ischemic stroke. Furthermore, we observed significant disregulation of multiple *PDE4D* isoforms in affected individuals. We propose that this gene is involved in the pathogenesis of stroke through atherosclerosis. *PDE4D* is expressed in cell types important in atherosclerosis and regulates a second messenger that has a central role in processes important in the pathogenesis of atherosclerosis. Perhaps inhibition of *PDE4D* in general, or of one or more isoforms specifically, by a small-molecule drug might decrease the risk of stroke in those who are predisposed by genotype at *PDE4D*.

METHODS

Subjects. We recruited individuals with stroke and carried out phenotypic subclassification as previously described[17]. The study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. We obtained informed consent from all affected individuals and their relatives whose DNA samples were used in the analyses. All personal identifiers associated with medical information and blood samples were encrypted with a third party encryption system by the Data Protection Commission[36]. The phenotypes of participating affected individuals were redetermined by neurologists examining the clinical and radiological records, and those affected with ischemic stroke or TIA were subcategorized according to the TOAST research criteria[4]. We used a cutoff of 70% stenosis as the criterion for carotid stroke.

Identification of DNA polymorphisms. We identified new polymorphic repeats (dinucleotide or trinucleotide repeats) with the Sputnik program. We subtracted the smaller allele of CEPH sample 1347-02 (CEPH genomics repository) from the alleles of the microsatellites and used it as a reference. We detected SNPs by sequencing exonic and intronic regions from affected individuals and controls by PCR. We also detected public polymorphisms by BLAST search of the US National Center for Biotechnology Information's SNP database. We genotyped SNPs using a method for detecting SNPs with fluorescent polarization template-directed dye-terminator incorporation (SNP-FP-TDI assay; ref. 37).

Statistical analysis. For single-marker association studies, we used Fisher's exact test to calculate two-sided *P* values for each individual allele. All *P* values are unadjusted for multiple comparisons unless specifically indicated. We present allelic rather than carrier frequencies for microsatellites, SNPs and haplotypes. To minimize any bias due to the relatedness of the affected individuals who were recruited as families for the linkage analysis, we eliminated first- and second-degree relatives from the list of affected individuals. We also repeated the test for association, correcting for any remaining relatedness among the affected individuals by extending a variance adjustment procedure described previously[38] for sibships to apply to general familial relationships, and present both adjusted and unadjusted *P* values for comparison. The differences are generally very small, as expected. To assess the significance of single-marker association corrected for multiple testing, we carried out a randomization test using the same genotype data. We randomized the cohorts of affected individuals and controls and redid the association analysis. This procedure was repeated up to 500,000 times, and the *P* value we present is the fraction of replications that produced a *P* value for some marker allele that is lower than or equal to the *P* value we observed using the original affected and control cohorts.

For both single-marker and haplotype analyses, we calculated relative risk (RR) and population attributable risk assuming a multiplicative model (haplotype relative risk model; refs. 18,19) in which the risks of the two alleles or haplotypes a person carries multiply. For example, if RR is the risk of allele A relative to allele a, then the risk of an AA homozygote will be RR times that of an Aa heterozygote and RR^2 times that of an aa homozygote. The multiplicative model simplifies analysis and computations because haplotypes are independent, meaning they are in Hardy-Weinberg equilibrium in the affected population as well as in the control population. As a consequence, haplotype counts of the affected individuals and controls each have multinomial distributions, but with different haplotype frequencies under the alternative hypothesis. Specifically, for two haplotypes h_i and h_j, $\text{risk}(h_i)/\text{risk}(h_j) = (f_i/p_i)/(f_j/p_j)$, where *f* and *p* denote frequencies in the affected population and in the control population, respectively. Although there is some power loss if the true model is not multiplicative, the loss tends to be mild except in extreme cases. Most importantly, *P* values are always valid because they are computed with respect to the null hypothesis.

In general, haplotype frequencies are estimated by maximum likelihood and tests of differences between affected individuals and controls are carried out using a generalized likelihood ratio test[39]. We used our haplotype analysis program, called NEMO (which stands for 'nested models'; see **Supplementary Note** online for more details), to calculate all the haplotype results presented. To handle uncertainties with phase and missing genotypes, we did not use the common two-step approach to association tests, in which haplotype counts are first estimated, possibly with the use of the EM algorithm[40], and tests are then carried out, treating the estimated counts as though they are true counts. This method can be problematic and may require randomization to properly evaluate statistical significance. Instead, with NEMO, maximum likelihood estimates, likelihood ratios and *P* values are computed with the aid of the EM algorithm directly for the observed data; hence, loss of information due to uncertainty with phase and missing genotypes is automatically captured by the likelihood ratios. Even so, how much information is retained or lost may be of interest; **Supplementary Note** online describes such a measure that is natural under the likelihood framework.

For a fixed set of markers, the simplest tests we did (with results presented in **Supplementary Table 6** online) compare one selected haplotype against all the others. Call the selected haplotype h_1 and the others be $h_2, \ldots h_k$. Let $p_1, \ldots p_k$ denote the population frequencies of the haplotypes in the controls, and let $f_1, \ldots f_k$ denote the population frequencies of the haplotypes in the affected individuals. Under the null hypothesis, $f_i = p_i$ for all *i*. The alternative model we use for the test assumes $h_2, \ldots h_k$ to have the same risk but h_1 has a different risk. This implies that p_1 can be different from f_1, but $f_i/(f_2 + \ldots + f_k) = p_i/(p_2 + \ldots + p_k) = \beta_i$ for $i = 2, \ldots k$. Denoting f_1/p_1 with *r*, and noting that $\beta_2 + \ldots + \beta_k = 1$, the test statistic based on generalized likelihood ratios is

$$\Lambda = 2\left[l(\hat{r}, \hat{p}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{k-1}) - l(1, \hat{p}_1, \tilde{\beta}_2, \ldots, \tilde{\beta}_{k-1})\right]$$

where *l* denotes $\log_e$ likelihood and ˜ and ^ denote maximum likelihood estimates under the null hypothesis and alternative hypothesis, respectively. Λ has asymptotically a χ^2 distribution with 1 degree of freedom under the null hypothesis, and it was used to compute *P* values presented in **Supplementary Table 6** online. The tests presented in **Figure 4** have slightly more complicated null and alternative hypotheses. For the results in **Figure 4a**, let h_1 be G0, h_2 be GX and h_3 be AX. When comparing G0 with GX (the test that gives estimated RR = 1.46 and *P* = 0.0002), the null hypothesis assumes G0 and GX have the same risk but AX has a different risk. The alternative hypothesis allows all three haplotype groups to have different risks. This implies that, under the null hypothesis, there is a constraint that $f_1/p_1 = f_2/p_2$, or $w = (f_1/p_1)/(f_2/p_2) = 1$. The test statistic based on generalized likelihood ratios is

$$\Lambda = 2\left[l(\hat{p}_1, \hat{f}_1, \hat{p}_2, \hat{w}) - l(\tilde{p}_1, \tilde{f}_1, \tilde{p}_1, 1)\right]$$

which again has asymptotically a χ^2 distribution with 1 degree of freedom under the null hypothesis. There is actually an extra complication to the test when h_2 and h_3 are composite haplotypes. That is handled in a natural manner under the nested models framework with details given in **Supplementary Note** online. Other tests presented in **Figure 4** were similarly carried out.

We calculated LD between pairs of SNPs using the standard definition of D' (ref. 41) and R^2 (ref. 42). Using NEMO, frequencies of the two marker allele combinations are estimated by maximum likelihood, and deviation from linkage equilibrium is evaluated by a likelihood ratio test. We extended the definitions of D' and R^2 to include microsatellites by averaging over the values for all possible allele combinations the two markers weighted by the marginal allele probabilities. When plotting all marker combinations to elucidate the LD structure in a particular region, we plotted D' in the upper left corner and the *P* value in the lower right corner. In the LD plots we present, the markers are plotted equidistantly rather than according to their physical positions.

Enquiries regarding information and accessibility of the haplotype analysis program NEMO should be addressed to A.K. (augustine.kong@decode.is) or D.G. (daniel.gudbjartsson@decode.is).

Expression analysis using quantitative reverse transcriptase PCR. We isolated total RNA from EBV-transformed B-cell cultures according to the manual using the TRIZOL reagent provided by GibcoBRL. We used the RNeasy mini Qiagen kit with on-column DNA digestion to clean RNA. We assessed the quality and quantity of RNA using 2100 Agilent Bioanalyser. We prepared cDNA from total RNA using random hexamers with TaqMan Reverse Transcription Reagents kit from Applied Biosystems (N808-0234). We used Primer Express 2.0 and Oligo 6 software to make cDNA-specific primers and probes for *PDE4D* and *PDE4D* isoforms. We obtained *GAPD* 'Assay-On-Demand' from Applied Biosystems and used it as a housekeeping gene. We tested PDE assays and optimized them for 384-well high-throughput expression analysis on ABI 7900 Instrument. We used a final concentration of 200 nM probes, 900 nM primers and 2 ng μl^{-1} cDNA in a 10-μl reaction volume. We processed each plate twice and calculated an average for each sample. We used the ABI7900 instrument to calculate CT (Threshold Cycle) values. We calculated quantity estimates using the formula $2^{-\Delta CT}$ where ΔCT represents the difference in CT values for target and housekeeping assays. We eliminated from our analyses any samples whose duplicates differed by more than 1 ΔCT.

URLs. The American Heart Association can be found at http://www.americanheart.org/. The Sputnik program can be found at http://espressosoftware.com/pages/sputnik.jsp. The US National Center for Biotechnology Information's SNP database is found at http://www.ncbi.nlm.nih.gov/SNP/index.html.

GenBank accession numbers. *PDE4D7*, AY245866; *PDE4D9*, AY245867.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the affected individuals and their families whose contribution made this study possible and the nurses at the Icelandic Heart Association, staff at the deCODE core facilities and V. Brophy and S. Cheng for their valuable contribution to this work. Enquiries regarding information and accessibility of the haplotype analysis program NEMO should be addressed to A.K. (augustine.kong@decode.is) or D.G. (daniel.gudbjartsson@decode.is).

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Received 1 July; accepted 27 August 2003
Published online at http://www.nature.com/naturegenetics/

1. Sacco, R.L. *et al.* American Heart Association Prevention Conference. IV. Prevention and Rehabilitation of Stroke. Risk factors. *Stroke* **28**, 1507–1517 (1997).
2. Bonita, R. Epidemiology of stroke. *Lancet* **339**, 342–344 (1992).
3. Caplan, L.R. *Caplan's Stroke: A Clinical Approach.* (Butterworth-Heinemann, Boston, 2000).
4. Adams, H.P.Jr. *et al.* Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* **24**, 35–41 (1993).
5. Fisher, C.M. Lacunar strokes and infarcts: a review. *Neurology* **32**, 871–876 (1982).
6. Alberts, M.J. *Genetics of Cerebrovascular Disease* (Futura, New York, 1999).
7. Hassan, A. & Markus, H.S. Genetics and ischaemic stroke. *Brain* **123**, 1784–1812 (2000).
8. Leys, D. *et al.* Stroke prevention: management of modifiable vascular risk factors. *J. Neurol.* **249**, 507–517 (2002).
9. Brass, L.M. & Alberts, M.J. The genetics of cerebrovascular disease. *Baillieres Clin. Neurol.* **4**, 221–245 (1995).
10. Tournier-Lasserve, E. *et al.* Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy maps to chromosome 19q12. *Nat. Genet.* **3**, 256–259 (1993).
11. Joutel, A. *et al. Notch3* mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature* **383**, 707–710 (1996).
12. Palsdottir, A. *et al.* Mutation in cystatin C gene causes hereditary brain haemorrhage. *Lancet* **2**, 603–604 (1988).
13. Levy, E. *et al.* Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type. *Science* **248**, 1124–1126 (1990).
14. Gunel, M., Awad, I.A., Anson, J. & Lifton, R.P. Mapping a gene causing cerebral cavernous malformation to 7q11.2–q21. *Proc. Natl. Acad. Sci. USA* **92**, 6620–6624 (1995).
15. Laberge-le Couteulx, S. *et al.* Truncating mutations in *CCM1*, encoding KRIT1, cause hereditary cavernous angiomas. *Nat. Genet.* **23**, 189–193 (1999).
16. Sahoo, T. *et al.* Mutations in the gene encoding KRIT1, a Krev-1/rap1a binding protein, cause cerebral cavernous malformations (CCM1). *Hum. Mol. Genet.* **8**, 2325–2333 (1999).
17. Gretarsdottir, S. *et al.* Localization of a susceptibility gene for common forms of stroke to 5q12. *Am. J. Hum. Genet.* **70**, 593–603 (2002).
18. Terwilliger, J.D. & Ott, J. A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* **42**, 337–346 (1992).
19. Falk, C.T. & Rubinstein, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann. Hum. Genet.* **51** (Pt 3), 227–233 (1987).
20. Houslay, M.D. & Adams, D.R. PDE4 cAMP phosphodiesterases: modular enzymes that orchestrate signalling cross-talk, desensitization and compartmentalization. *Biochem. J.* **370**, 1–18 (2003).
21. Liu, H. & Maurice, D.H. Phosphorylation-mediated activation and translocation of the cyclic AMP- specific phosphodiesterase PDE4D3 by cyclic AMP-dependent protein kinase and mitogen-activated protein kinases. A potential mechanism allowing for the coordinated regulation of PDE4D activity and targeting. *J. Biol. Chem.* **274**, 10557–10565 (1999).
22. Liu, H. *et al.* Expression of phosphodiesterase 4D (PDE4D) is regulated by both the cyclic AMP-dependent protein kinase and mitogen-activated protein kinase signaling pathways. A potential mechanism allowing for the coordinated regulation of PDE4D activity and expression in cells. *J. Biol. Chem.* **275**, 26615–26624 (2000).
23. Baillie, G., MacKenzie, S.J. & Houslay, M.D. Phorbol 12-myristate 13-acetate triggers the protein kinase A-mediated phosphorylation and activation of the PDE4D5 cAMP phosphodiesterase in human aortic smooth muscle cells through a route involving extracellular signal regulated kinase (ERK). *Mol. Pharmacol.* **60**, 1100–1111 (2001).
24. Jin, S.L. & Conti, M. Induction of the cyclic nucleotide phosphodiesterase PDE4B is essential for LPS-activated TNF-α responses. *Proc. Natl. Acad. Sci. USA* **99**, 7628–7633 (2002).
25. Landells, L.J. *et al.* Identification and quantification of phosphodiesterase 4 subtypes in CD4 and CD8 lymphocytes from healthy and asthmatic subjects. *Br. J. Pharmacol.* **133**, 722–729 (2001).
26. Fukumoto, S. *et al.* Distinct role of cAMP and cGMP in the cell cycle control of vascular smooth muscle cells: cGMP delays cell cycle transition through suppression of cyclin D1 and cyclin-dependent kinase 4 activation. *Circ. Res.* **85**, 985–991 (1999).
27. Ogawa, S. *et al.* Hypoxia-induced increased permeability of endothelial monolayers occurs through lowering of cellular cAMP levels. *Am. J. Physiol.* **262**, C546–C554 (1992).
28. Stelzner, T.J., Weil, J.V. & O'Brien, R.F. Role of cyclic adenosine monophosphate in the induction of endothelial barrier properties. *J. Cell Physiol.* **139**, 157–166 (1989).
29. Pan, X., Arauz, E., Krzanowski, J.J., Fitzpatrick, D.F. & Polson, J.B. Synergistic interactions between selective pharmacological inhibitors of phosphodiesterase isozyme families PDE III and PDE IV to attenuate proliferation of rat vascular smooth muscle cells. *Biochem. Pharmacol.* **48**, 827–835 (1994).
30. Palmer, D., Tsoi, K. & Maurice, D.H. Synergistic inhibition of vascular smooth muscle cell migration by phosphodiesterase 3 and phosphodiesterase 4 inhibitors. *Circ. Res.* **82**, 852–861 (1998).
31. Indolfi, C. *et al.* Activation of cAMP-PKA signaling *in vivo* inhibits smooth muscle cell proliferation induced by vascular injury. *Nat. Med.* **3**, 775–779 (1997).
32. Indolfi, C. *et al.* 8-chloro-cAMP inhibits smooth muscle cell proliferation *in vitro* and neointima formation induced by balloon injury *in vivo*. *J. Am. Coll. Cardiol.* **36**, 288–293 (2000).
33. Tilley, S.L., Coffman, T.M. & Koller, B.H. Mixed messages: modulation of inflammation and immune responses by prostaglandins and thromboxanes. *J. Clin. Invest.* **108**, 15–23 (2001).
34. Vicini, E. & Conti, M. Characterization of an intronic promoter of a cyclic adenosine 3′,5′-monophosphate (cAMP)-specific phosphodiesterase gene that confers hormone and cAMP inducibility. *Mol. Endocrinol.* **11**, 839–850 (1997).
35. Le Jeune, I.R., Shepherd, M., Van Heeke, G., Houslay, M.D. & Hall, I.P. Cyclic AMP-dependent transcriptional up-regulation of phosphodiesterase 4D5 in human airway smooth muscle cells. Identification and characterization of a novel PDE4D5 promoter. *J. Biol. Chem.* **277**, 35980–35989 (2002).
36. Gulcher, J.R., Kristjansson, K., Gudbjartsson, H. & Stefansson, K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur. J. Hum. Genet.* **8**, 739–742 (2000).
37. Chen, X., Zehnbauer, B., Gnirke, A. & Kwok, P.Y. Fluorescence energy transfer detection as a homogeneous DNA diagnostic method. *Proc. Natl. Acad. Sci. USA* **94**, 10756–10761 (1997).
38. Risch, N. & Teng, J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. *Genome Res.* **8**, 1273–1288 (1998).
39. Rice, J.A. Generalized likelihood ratio tests. in *Mathematical Statistics and Data Analysis* 308–310 (International Thomson Publishing, Belmont, California, 1995).
40. Dempster, A.P., Laird, N.M. & Rubin, D.B. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. R. Stat. Soc. Ser. B Stat. Med.* **39**, 1–38 (1971).
41. Lewontin, R. The interaction of selection and linkage I. General considerations: Heterotic models. *Genetics* **49**, 49–67 (1964).
42. Hill, W.G. & Robertson, A. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **22**, 226–231 (1968).

Original Paper

Hum Hered 1992;42:337–346

A Haplotype-Based 'Haplotype Relative Risk' Approach to Detecting Allelic Associations

Joseph D. Terwilliger
Jurg Ott

Department of Genetics and Development,
Columbia University,
New York, N.Y., USA

Key Words
Linkage disequilibrium
Haplotype relative risk
Allelic association,
Chi-square tests

Abstract
A novel variation of the Haplotype Relative Risk (HRR) of Rubinstein et al. [Hum Immunol 1981;3:384] is proposed, in order to glean increased information about linkage disequilibrium or allelic associations by analyzing haplotype-based data rather than genotypic data. It is shown that statistical tests based on our design give much higher power than those based on the original HRR approach. Several additional nonparametric tests based on the same data are analyzed, and power is computed for each of them. Further, parametric likelihood methods are applied to testing linkage equilibrium, and estimating δ, the coefficient of linkage disequilibrium, from the same data.

Introduction

Allelic associations between etiologically unrelated traits were originally detected in humans through observations at the genotypic level. In the 1950s, it was noticed that in individuals with certain diseases there were significant excesses of certain blood groups. Aird et al. [1, 2] demonstrated the presence of a significant association between blood group A and stomach cancer, and between blood group O and peptic ulcer, while Pike and Dickens [3] found such an association between blood group O and toxemia of pregnancy, and McConnell et al. [4] studied associations between blood groups and carcinoma of the lung. Woolf [5] then proposed his Relative Risk statistic to compare the incidence rates in given blood groups in a case control type of study, in which one would collect a sample of people with the disease and compare the observed frequency of the 'risk allele' with its frequency in a separate sample of healthy individuals (or population frequency, if known).

One problem with this method is that there is no way of knowing whether a significant re-

Joseph D. Terwilliger
722 West 168th St. Box 58
New York, NY 10032 (USA)

iff B, Gordon L, Haendel S, th LK, Carrano AV: The somal constitution of human selected for motility. Fertil 986;46:686–690.
TA, Martin RH, Hoar DI: nent of the sephadex tech- or selection of X chromo- earing human sperm by anal- sperm chromosomes, deoxy- leic acid and Y-bodies. Fertil 989;52:829–835.
iff BF, Gordon LA, Haendel r S, Moore DI, Gledhill BL: omosome ratios determined otypic analysis in albumin- l human sperm. Fertil Steril :678–685.
CU, Yanagimachi R: Sperm some analysis as a new sys- test human X- and Y-sperm ion. Gamete Res 1987;17: 3.
ichi Y, Mikamo K: An im- efficient method for analys- nan sperm chromosomes us- ia-free hamster ova. Am J enet 1986;38:724–740.
do C, Benet J, Genesca A, o J, Caballin MR, Miro R, e J: Human sperm chromo- Hum Reprod 1988;3:133–

RH. Sex ratio among sperm m J Hum Genet 1990;47:349–

sult is biologically meaningful or just a consequence of having the case and control samples taken from different genetic populations in which the frequency of the risk allele is different and therefore, no real association exists. To attempt to circumvent this problem, Rubinstein et al. [6] proposed the Haplotype Relative Risk (HRR) statistic, based on earlier work of C.A.B. Smith, to ensure that the control and disease samples were well-matched, from the same population, so that any observed association would have to be due to a real allelic association of some sort. This experimental design has also been used in the haplotype frequency difference statistic of Seuchter et al. [7].

Experimental Design

- H = Marker allele with which disequilibrium is hypothesized.
- $\overline{H}$ = Any allele other than H at the marker locus.
- δ = Gametic linkage desiquilibrium coefficient; = P(AB gamete) – P(A)P(B) (A at one locus B at the other).
- Θ = Recombination fraction between marker and disease loci.
- p = Gene frequency of the disease allele.
- q = Gene frequency of the H allele.
- n = Sample size.

In order to be sure one has matched control and disease samples, Rubinstein et al. [6] proposed using data from nuclear families with one affected offspring to test for deviations from linkage equilibrium. They recommended using the affected offspring's genotype (made up of alleles transmitted from parents to the affected child) at a marker locus as the 'case' sample, and an artificial genotype made up of the alleles not transmitted to the child from its parents as the 'control' sample in an association test. Then they used such data to test whether the H allele was present equally frequently in diseased individuals' genotypes, and the nontransmitted control genotypes. For example, in a family with unaffected parents with genotypes G/H and I/J at the marker locus, and an affected child with marker genotype H/I, the transmitted genotype would be H/I, and the artificial nontransmitted genotype would be G/J. Since they were only interested in whether H was present or absent from the genotypes, in this example we have H transmitted, and $\overline{H}$ not transmitted (genotype G/J does not contain H). For every such nuclear family there would be one such observation. One can then tabulate such observations in the form of table 1. The example family above would fall in cell B. Ott [8] demonstrated that under the null hypothesis of $\delta = 0$, the transmitted and nontransmitted

Table 1. Data collected in a haplotype relative risk study (either HHRR, or GHRR)

Transmitted	Not transmitted		Total
	H	$\overline{H}$	
H	A	B	W
$\overline{H}$	C	D	X
Total	Y	Z	N

In the 2×2 table shown here, each cell corresponds to one parent. In the HHRR, each parent transmits one allele, and not the other, and can thus be classified by which allele was, and which was not transmitted to the affected offspring. In the GHRR, each set of parents has 4 alleles, 2 of which are transmitted to the affected child, and 2 which are not. If the child contains 1 or 2 H alleles, we say H was transmitted, and if there is an H allele in the remaining 2 alleles, we say that H was not transmitted. Thus, each family either transmits H or $\overline{H}$, and has either H or $\overline{H}$ among the nontransmitted alleles, and can therefore also be characterized by one cell of this table.

Table 2. Haplotype relative risk

	H	$\overline{H}$	Total
Transmitted	W	X	N
Not transmitted	Y	H	N
Total	W+Y	X+Z	2N

The data in this table are taken directly from the marginals of table 1, and represent the form of the originally proposed GHRR statistic. This table, of course, can be filled with either haplotype- or genotype-based data. All variable names are the same as in table 1.

a haplotype relative risk
RR)

ansmitted	Total
$\overline{H}$	
B	W
D	X
Z	N

1 here, each cell corre-
ne HHRR, each parent
the other, and can thus
was, and which was not
offspring. In the GHRR,
les, 2 of which are trans-
and 2 which are not. If
eles, we say H was trans-
allele in the remaining 2
transmitted. Thus, each
$\overline{H}$, and has either H or $\overline{H}$
lleles, and can therefore
cell of this table.

ve risk

	$\overline{H}$	Total
	X	N
	H	N
	X+Z	2N

taken directly from the
present the form of the
statistic. This table, of
her haplotype- or geno-
names are the same as

sent from the genotypes,
transmitted, and $\overline{H}$ not
s not contain H). For ev-
would be one such obser-
such observations in the
amily above would fall in
that under the null hy-
tted and nontransmitted

Based HRR

alleles are independently associated, and thus we can treat our transmitted and nontransmitted samples independently and represent them in the form of table 2 (marginals of table 1). Then a standard χ^2 test of independence on this table can be shown to be a valid χ^2 test of the hypothesis $\delta = 0$. This is the test proposed by Rubinstein et al. [6] to guarantee the control and disease samples are genetically well-matched.

As is shown below, the statistical method of Rubinstein et al. [6] does not take advantage of all the information present in the data. Their method lumps H/H homozygotes and H/$\overline{H}$ heterozygotes together as H genotypes. However, since under the null hypothesis the two parental genotypes are independent, it is possible to treat each parent as an independent observation, and merely look at the fate of each parental marker allele. So, in the example family above, there would be one observation of H transmitted, G not transmitted, and one observation of I transmitted, J not transmitted, which in table 1 (now referring to alleles, not genotypes), would contribute one observation to cell B, and one observation to cell D. Again, for theoretical reasons given by Ott [8], transmitted and nontransmitted alleles are independent for each other, and can be collapsed, as in the Rubinstein case, into table 2, in which the example family would contribute one observation to cell W, one to cell X, and two to cell Z, the marginal values of table 1. We are thus using more of the information present in the family, obtaining twice as many observations from the same amount of data.

Recessive Disease

Haplotype-Based versus Genotype-Based HRR χ^2 Tests

We first compared the power of our haplotype-based HRR (HHRR) statistic with the genotype-based HRR (GHRR) of Rubinstein et al. [6]. The test we applied to each data set is essentially a χ^2 test of independence on table 2 for the haplotype-based data (HHRR test), and for the equivalent genotype-based table (GHRR test) in which discrimination is between genotypes with no H allele, and those with at least one (possibly two). Power calculations were performed for each test, assuming a recessive disease with no phenocopies (penetrance is irrelevant to the calculations, according to Ott [8]), by analytically computing the probability of a significant χ^2 test result ($\chi^2 > 3.84$ at the 0.05 level) for different combinations of δ/p (δ and p are completely confounded according to Ott [8]), q, and Θ. Power curves for these two tests (n = 100 families, q = 0.5) are given in figure 1 for varying true values of Θ and δ/p. In all the numerical cases we considered, the HHRR test was more powerful than the GHRR approach of Rubinstein et al. [6]. This is intuitively satisfying, since the HHRR approach discriminates between H/H homozygotes and H/$\overline{H}$ heterozygotes, while the GHRR does not. Thus, our approach uses all of the information in the data, where the traditional GHRR does not.

The test of independence on table 2 is a test of E[W] = E[Y]. However, W and Y are obtained from the marginals of table 1. So, when we are testing E[W] = E[Y], we are essentially testing E[A + B] = E[A + C], which is the same as E[B] = E[C]. Clearly this is expected under the null hypothesis of no disequilibrium. Using the data from table 1, the HHRR χ^2 is computed as

$$\frac{2N(B-C)^2}{(2A+B+C)(N-2A-B-C)}$$

$$= \frac{2N(WZ-XY)^2}{(W+X)(W+Y)(X+Z)(Y+Z)},$$

the standard χ^2 test of independence on a 2×2 table. This is a valid χ^2 test, of the form $(B-C)^2/\mathrm{Var}[B-C]$, since $\mathrm{Var}[B-C] = 2Nq(1-q)$, which is estimated by $2N[(2A+B+C)/(2N)][1-(2A+B+C)/(2N)]$. The power is shown graphically in figure 2 for n = 50 families (for comparison with other haplotype-based tests below).

McNemar Tests

Since our null hypothesis is B = C in a paired sampling (transmitted allele, nontransmitted allele) test, one's first intuition might

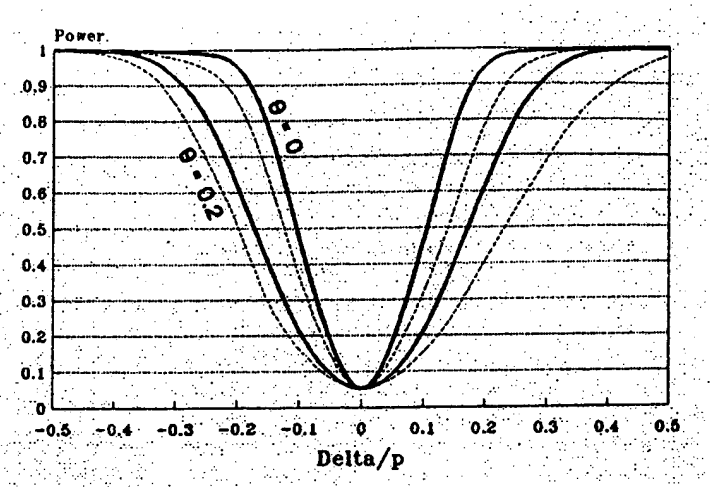


Fig. 1. Power curves (analytically computed) for χ^2 tests based on the haplotype- (—) and genotype-based (----) HRR designs (100 families), for $q = 0.5$. If $p = 0.5$, then all values of δ/p shown are possible. For other values of p, different restrictions apply, but have no effect on the power curve. The upper two lines are for the power of the test when $\Theta = 0$, and the lower set of two lines correspond to $\Theta = 0.20$. Note that the haplotype-based design yields higher power for all true values of Θ and δ/p.

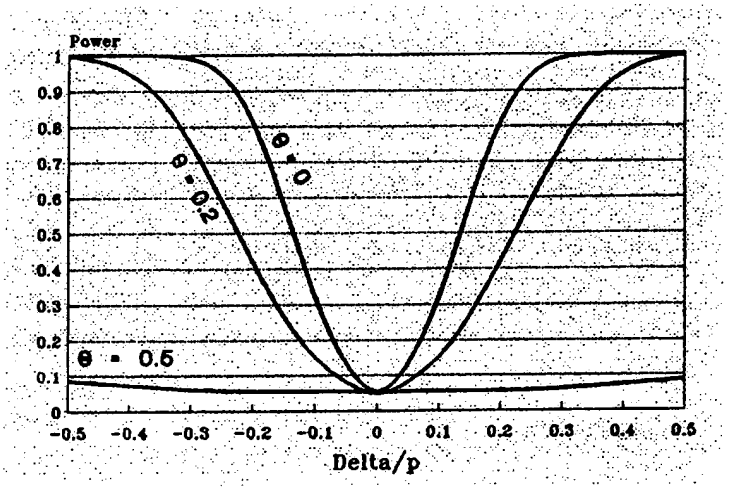
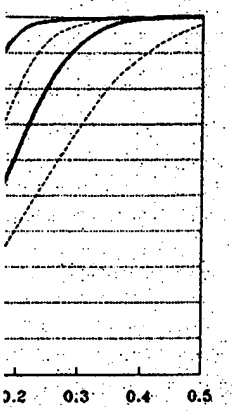


Fig. 2. Power curves (analytically computed) for the HHRR test (50 families) for $q = 0.5$, with $\Theta = 0$ (upper curve), 0.2 (middle curve) and 0.5 (lower curve).

be to apply a McNemar test, $(B-C)^2/(B+C)$. In order for this to also be a valid χ^2 test, $(B+C)$ would have to be an estimate of the variance of $(B-C)$, which we already have shown to be $2Nq(1-q)$. Our HHRR χ^2 test uses all the data to estimate q, including the information from homozygous individuals, while in the McNemar test, all homozygotes are ignored, and the variance is estimated as $(B+C)$. Clearly $E[C] = E[B] = Nq(1-q)$ under the null hypothesis ($\delta = 0$), so $(B+C)$ then estimates $2Nq(1-q)$. However, in every numerical case we considered, this test was less powerful than the HHRR test, as shown in figure 3, due to the fact that the HHRR uses all of the data to estimate the variance, while the McNemar uses only the information from heterozygous parents.

Independence Tests

An interesting result of Ott [8] is that transmitted and nontransmitted alleles are inde-



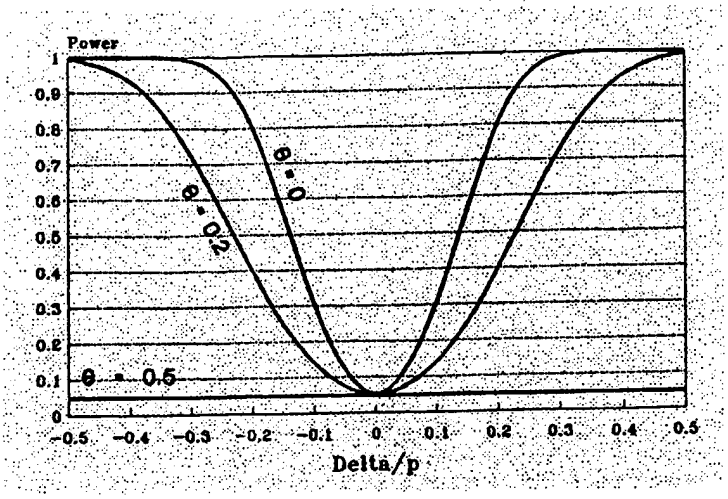
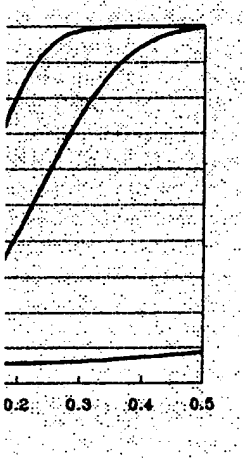


Fig. 3. Power curves (analytically computed) for the haplotype-based McNemar (HMCN) test (50 families) for q = 0.5, with Θ = 0 (upper curve), 0.2 (middle curve), and 0.5 (lower curve).



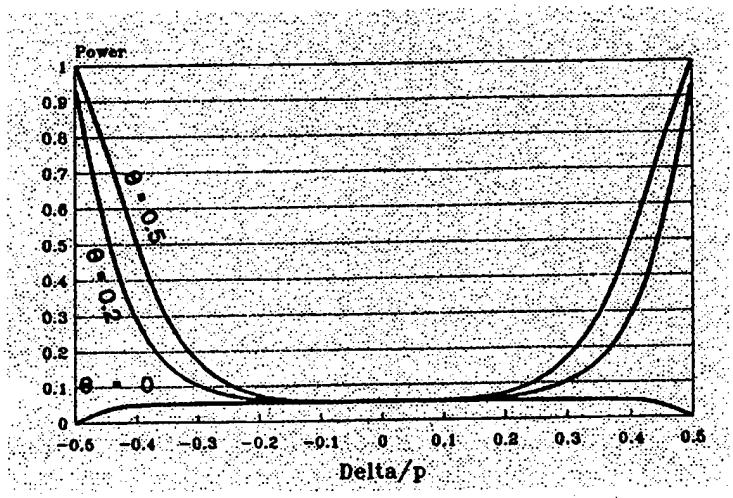


Fig. 4. Power curves (analytically computed) for the haplotype-based independence test (HIND) for 50 families, q = 0.5, and Θ = 0 (lower curve), 0.2 (middle curve), and 0.5 (upper curve).

ver, in every numer-
his test was less pow-
t, as shown in figure
e HHRR uses all of
variance, while the
formation from het-

Ott [8] is that trans-
ed alleles are inde-

pendent when δ = 0 or when Θ = 0. In light of this, one could use an independence test on table 1 as a test of δ = 0, though clearly when Θ is close to 0, this test should not be useful. This test is just that (AD–BC) = 0. Therefore, the test should be $(AD–BC)^2/Var(AD–BC)$, which is the standard χ^2 test of independence on a 2 × 2 table, $N(AD–BC)^2/(WXYZ)$. Power was analytically computed for this test, under the recessive model, for various true values of q, δ/p, and Θ, which are graphically presented in figure 4. In this test, the power increases as Θ increases, just the opposite behavior from the HHRR and McNemar tests. This test may thus be a useful way to use such nuclear family data to test δ = 0 when Θ is known to be quite large, since when Θ = 0.5, the HHRR tends to 0 [8].

This independence test, however, fails to impose the restriction that the frequency of the H allele be equal in both the transmitted and nontransmitted samples. To include this

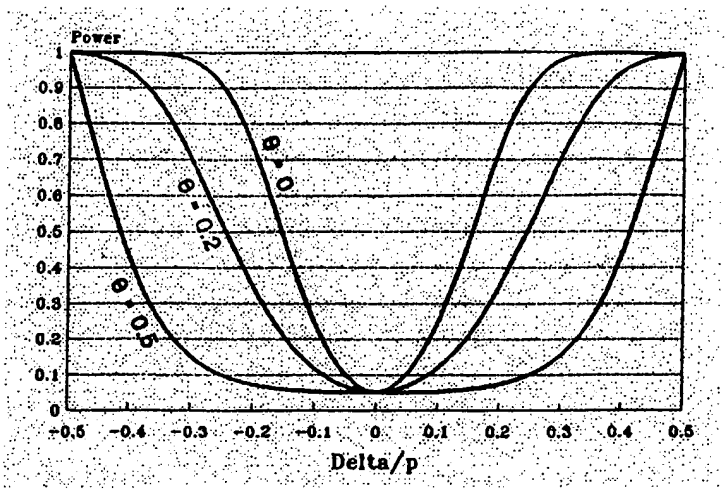


Fig. 5. Power curves (analytically computed) for the test of fit to the expected multinominal proportions (HIID) of haplotype-based data for 50 families, q = 0.5, and Θ = 0 (upper curve), 0.2 (middle curve), and 0.5 (lower curve).

information, one could test the fit of the counts of A, B, C, and D to their expected multinominal proportions (each observation is clearly independent) as follows: $\Sigma(O-E)^2/E$, which is equal to

$$\frac{(A-N\hat{q}^2)^2}{N\hat{q}^2}+\frac{[B-N\hat{q}(1-\hat{q})]^2}{N\hat{q}(1-\hat{q})}+\frac{[C-N\hat{q}(1-\hat{q})]^2}{N\hat{q}(1-\hat{q})}$$

$$+\frac{[D-N(1-\hat{q})^2]^2}{N(1-\hat{q})^2},\ \text{where}\ \hat{q}=\frac{2A+B+C}{2N}.$$

This test follows a χ^2 distribution with 2 df, since we had 4 cell counts, but fixed the sum A + B + C + D = N, and estimated q from the data. This test is very powerful over a large range of values of δ/p, q, and Θ, as shown in figure 5, and thus provides a useful general test for disequilibrium.

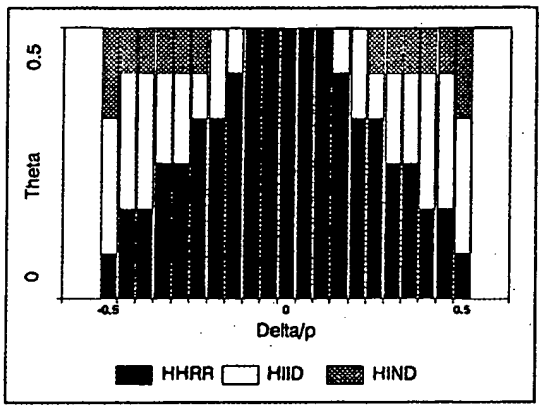
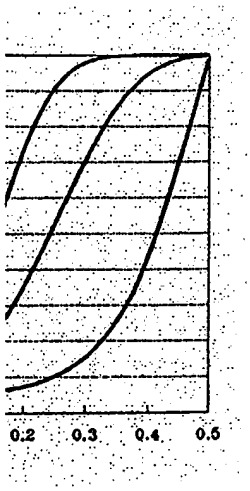
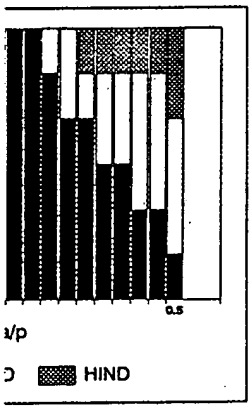


Fig. 6. Graph showing, for all possible values of Θ and δ/p, and fixed q = 0.5, which among three tests is the most powerful (50 families). The values of the power are not shown, but are given in fig. 2–5 (HMCN is never the most powerful).

Relative Power of Nonparametric Approaches

Each of the tests described above has different properties which make it useful. However, the question remains as to which test should be used in a given situation. To answer that question, for each combination of Θ, δ/p, and q, we determined which test gave maximal power for a sample size of 50 families. The results are presented graphically in figure 6. In this figure, for fixed q, we considered all possible combinations of δ/p and Θ, and determined which test gave maximal power (analytically computed). Then for each point (δ/p, Θ) the most powerful test is indicated. To see ex-





r all possible values of Θ
hich among three tests is
lies). The values of the
given in fig. 2-5 (HMCN

ich test gave maximal
of 50 families. The re-
hically in figure 6. In
e considered all pos-
p and Θ, and deter-
ximal power (analyt-
or each point (δ/p, Θ)
indicated. To see ex-

actly what the power was, the reader is referred to the power curves already presented for each test. Some interesting patterns can be seen in this figure, but it should be used only in conjunction with the actual values of the power shown in figures 2–5, for often the difference is small between tests. However, over the most relevant ranges of δ/p and Θ, for all q, the HHRR test is the most powerful. In light of this, and the relative implausibility of strong disequilibrium when Θ is large, the HHRR test should be the general nonparametric test of choice, both for its power, and its simplicity.

Parametric Likelihood Ratio Tests

If one knows the model of the disease, one could do a parametric likelihood ratio test analysis, based on theoretical probabilities of each type of parent under a fixed model. Table 2 of Ott [8] provides such parametric values for the case of a recessive disease. The difficulty here is three fold. First, one needs to have an accurate parametric model for the disease, and compute the parametric probabilities of each cell of table 1. This process is very tedious (except for the recessive model described by Ott [8]), and depends heavily on the disease model. Secondly, one needs to maximize the likelihood of the data over all the parameters, Θ, (δ/p), and q, and then again maximize the likelihood, fixing $\delta = 0$. This would give us the following likelihood ratio: $L(\hat{\delta}/p, \hat{\Theta}, \hat{q})/L(\delta/p = 0, \hat{\Theta}, \hat{q})$. Normally, one can treat $2 \times \ln(LR)$ as a χ^2 random variable, with the number of degrees of freedom being the difference in free parameters in numerator and denominator of the likelihood ratio, which would appear to be 1 in this case. However, when $\delta = 0$, Θ disappears as a parameter, as shown by Ott [8]. When a parameter disappears under the null hypothesis, it is a degenerate situation, and so the statistic does not satisfy the criteria for χ^2. As the distribution is unclear, this test becomes very awkward to interpret, and presents a situation analogous to the degenerate likelihood ratio test for linkage in the presence of heterogeneity [9]. For this reason, combined with the enormous computer time involved, power was not calculated for this approach.

For general pedigree data (including nuclear families with multiple offspring), with a fixed-disease model, parametric likelihood ratio tests are tractable using any linkage analysis program, like ILINK of the LINKAGE package. One need only maximize the likelihood over Θ, q, p, and δ for the numerator, and again maximize the likelihood for the denominator over Θ, q, and p, fixing $\delta = 0$. This would then be a valid, and powerful general likelihood ratio test of $\delta = 0$, $2 \times \ln[L(\hat{\Theta}, \hat{\delta}, \hat{p}, \hat{q})/L(\hat{\Theta}, \delta = 0, \hat{p}, \hat{q})]$. It is important to remember that when using this method, the maximum likelihood estimates of the haplotype frequencies will reflect the sample frequency of the disease allele, which is not an accurate reflection of its population frequency. One must be sure to weight disease and control haplotypes. accordingly. For example, if our haplotype frequency estimates are $\hat{P}(Hd)$, $\hat{P}(\bar{H}d)$, $\hat{P}(HD)$, $\hat{P}(\bar{H}D)$, and we know the true gene frequency of the d allele, p_d, we can compute adjusted haplotype frequency estimates as

$$\tilde{P}(Hd) = \left(\frac{\hat{P}(Hd)}{\hat{P}(Hd) + \hat{P}(\bar{H}d)}\right)(p_d),$$

and so on. Similarly, if one wanted to estimate the coefficient of disequilibrium from such ILINK estimates, it would be necessary to use the adjusted estimates described above, yielding an adjusted estimate of

$$\tilde{\delta} = (\hat{\delta})\frac{p_d(1-p_d)}{\hat{p}_d(1-\hat{p}_d)},$$

where $\hat{\delta} = \hat{P}(Hd)\,\hat{P}(\bar{H}D) - \hat{P}(HD)\,\hat{P}(\bar{H}d)$, and $\hat{p}_d = \hat{P}(Hd) + \hat{P}(\bar{H}d)$.

An ad hoc method sometimes used in general pedigrees is to assume the absence of recombination, and determine the haplotypes of each founder, between marker and disease, as a way to insure the control (nondisease) haplotype are from the same genetic population as the disease haplotypes. This ad hoc approach has been applied, for example, in cystic fibrosis [10]. It assumes an absence of recombination, and its statistical properties are, in general, unclear, especially in cases where Θ is actually greater than zero. Another problem is that it is not always possible to uniquely and accurately determine all founder haplotypes. Censoring such indiscernible cases in some instances can be shown to lead to a statistical bias. In light of all of this, if one wants to use general pedigree data to test and quantify disequilibrium, the likelihood ratio test with ILINK described above is the test of choice, as it is more general and powerful, and has well-characterized statistical properties.

Nonrecessive Case

All of our results above were obtained for the case of a recessive disease. However, when other more complicated models prevail, the situation becomes unclear. While under any model we choose for the disease, the above tests are valid tests of $\delta = 0$ (since this implies no association between the disease and the marker locus), the effect on the power of our testing procedures is not so clear. When dealing with a recessive disease, a lot of additional information about linkage disequilibrium is obtained by looking at each parent separately, since each parent transmits a disease allele to the affected offspring, but the situation is less clear when there is a different model. For a dominant disease, with one affected parent, and one affected child, one can just consider the affected parent, and his or her transmitted and nontransmitted alleles, and base a test on the same procedure as above. The effect would be that there would be only one observation per family instead of two in the recessive case (where we know the parents to be heterozygous for the disease), and there is possible noise when the unaffected parent actually transmits the disease to the offspring, though this should be very rare.

In the case of dominant reduced-penetrance disease, in which neither parent is affected, clearly at least one parent must carry the disease-predisposing allele, though we cannot discern which one. In this situation, one parent will transmit the disease allele (in putative disequilibrium with the marker), and the other parent will transmit the normal allele. This adds noise to our system. One would expect the Rubinstein method to be less sensitive to this noise, since it doesn't distinguish between heterozygotes and homozygotes for the H allele.

Power calculations were approximated for this situation by simulation. A simplified model was considered in which one parent was forced to transmit the disease allele to the affected child, while the other parent was assumed to be homozygous unaffected (a reasonable assumption for small p). In this case, δ and p are no longer completely confounded, so we had to treat p, q, δ, and Θ as separate parameters. Then, 20,000 sets of 100 such nuclear families with 2 unaffected parents and one affected offspring were simulated under various assumptions on p, q, δ, and Θ. For each set of 100 families, the HHRR and GHRR were calculated. Then the number of significant results for each test at the 0.05 level ($\chi^2 \geq 3.84$) was counted to estimate the power of each test, which is graphed in figure 7. An interesting situation arises here, where the HHRR is much more powerful for negative values of δ, but for positive values of δ they are just about equal in power, with the GHRR being slightly

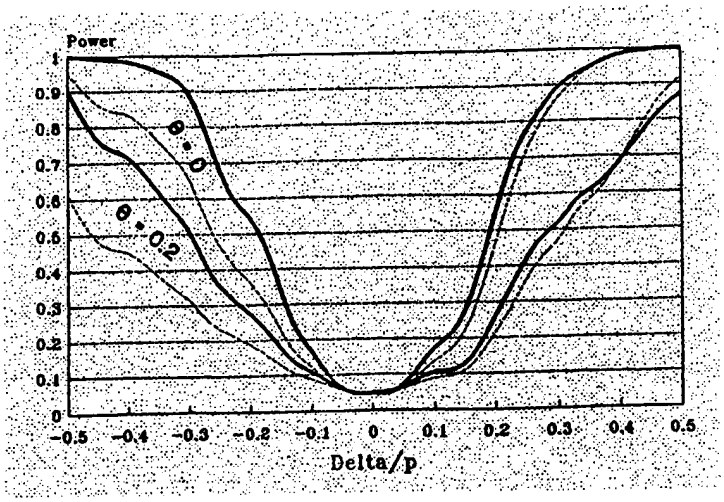


Fig. 7. Power curves (simulated) for the HHRR (——) and GHRR (----) tests with a dominant disease (reduced penetrance) and two unaffected parents, for $q = 0.5$, $p = 0.01$, and 100 families, based on 20,000 replicates. The upper curves represent $\Theta = 0$, and the lower curves $\Theta = 0.2$. In most cases, the HHRR is shown to be much more powerful than the GHRR.

more powerful for very extreme values of δ. The HHRR test is also more powerful than the other haplotype-based nonparametric tests over most of the reasonable sample space. The HHRR is more powerful than the GHRR in all recessive situations, dominant situations with $\delta < 0$, and about equally powerful with the GHRR in dominant situations with extremely positive δ. Further, the HHRR can take advantage of dominant situations with one affected parent, while the GHRR cannot. Therefore, we recommend using the HHRR as the nonparametric test of choice in general.

Discussion

When doing an association study, it is often difficult to find genetically well-matched cases and control samples. The HRR approach of using transmitted and nontransmitted alleles from the same parent as case and control samples ensures that they are genetically well-matched [11]. Further, the case and control samples are shown to be independent under the null hypothesis of $\delta = 0$. In light of this, HRR-type methods should be increasingly more important as geneticists try to map complex diseases, by looking for associations with candidate genes for example. In such a case, if the candidate gene is correct, Θ would be equal to 0, and these methods would achieve maximal power to detect the associations. Further, the built-in genetic control should provide a solution to the often difficult task of finding a valid control sample, and should allow people to have more faith in the validity of such association studies.

The approach presented here extracts further information about disequilibrium from the data used in the original GHRR approach, and thus presents a more powerful way to detect such associations in the absence of a parametric model. Given a parametric model, two likelihood-based methods were discussed as well. However, from the results of our power calculations, our HHRR seems to be the best general nonparametric test considered for detecting such associations with this experimental design over the most biologically plausible ranges of δ and Θ.

Acknowledgements

This work was supported by the National Institute of Mental Health (grant MH44292), the MacArthur Foundation Network I (Task Force on the Development of Analytic Strategies for Linkage Studies of Psychiatric Disorders), and the W.M. Keck Foundation. Further, the authors would like to acknowledge the contributions of Dr. Warren Ewens, whose comments are greatly appreciated.

References

1 Aird I, Bentall HH, Roberts JAF: The relationship between cancer of the stomach and the ABO blood groups. BMJ 1953;1:799–801.

2 Aird I, Bentall HH, Mehigan JA, Roberts JAF: The blood groups in relation to peptic ulceration and carcinoma of colon, rectum, breast, and bronchus. BMJ 1954;2:315–321.

3 Pike LA, Dickens AM: ABO blood groups and toxaemia of pregnancy. BMJ 1954;2:321–323.

4 McConnell RB, Clarke CA, Downton F: Blood groups in carcinoma of the lung. BMJ 1954;2:323–325.

5 Woolf B: On estimating the relation between blood and disease. Ann Hum Genet 1955;19:251–253.

6 Rubinstein P, Walker M, Carpenter C, Carrier C, Krassner J, Falk CT, Ginsburg F: Genetics of HLA disease associations. The use of the Haplotype Relative Risk (HRR) and the 'Haplo-Delta' (Dh) estimates in juvenile diabetes from three racial groups. Hum Immunol 1981;3:384.

7 Seuchter SA, Knapp M, Baur MP: Analysis of association in nuclear families; in Lynch HT, Tautu P (eds): Recent Progress in the Genetic Epidemiology of Cancer. Berlin, Springer, 1990, pp 89–94.

8 Ott J: Statistical properties of the Haplotype Relative Risk. Genet Epidemiol 1989;6:127–130.

9 Ott J: Analysis of Human Genetic Linkage. Baltimore, Johns Hopkins University Press, 1991.

10 Estivill X, Farrall M, Williamson R, Ferrari M, Seia M, Giunta AM, Novelli G, Potenza L, Dallapicolla B, Borgo G, Gasparini P, Pignatti PF, De Benedetti L, Vitale E, Devoto M, Romeo G: Linkage disequilibrium between cystic fibrosis and linked DNA polymorphisms in Italian families: A collaborative study. Am J Hum Genet 1988;43:23–28.

11 Falk CT, Rubinstein P: Haplotype Relative Risks: An easy way to construct a control sample for risk calculations. Ann Hum Genet 1987;51:227–233.

nature genetics

A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction

Anna Helgadottir[1], Andrei Manolescu[1], Agnar Helgason[1], Gudmar Thorleifsson[1], Unnur Thorsteinsdottir[1], Daniel F Gudbjartsson[1], Solveig Gretarsdottir[1], Kristinn P Magnusson[1], Gudmundur Gudmundsson[1], Andrew Hicks[1], Thorlakur Jonsson[1], Struan F A Grant[1], Jesus Sainz[1], Stephen J O'Brien[2], Sigurlaug Sveinbjornsdottir[3], Einar M Valdimarsson[3], Stefan E Matthiasson[3], Allan I Levey[4], Jerome L Abramson[4], Murdach P Reilly[5], Viola Vaccarino[4], Megan L Wolfe[5], Vilmundur Gudnason[6], Arshed A Quyyumi[4], Eric J Topol[7], Daniel J Rader[5], Gudmundur Thorgeirsson[3], Jeffrey R Gulcher[1], Hakon Hakonarson[1], Augustine Kong[1] & Kari Stefansson[1]

Variants of the gene *ALOX5AP* (also known as *FLAP*) encoding arachidonate 5-lipoxygenase activating protein are known to be associated with risk of myocardial infarction[1]. Here we show that a haplotype (HapK) spanning the *LTA4H* gene encoding leukotriene A4 hydrolase, a protein in the same biochemical pathway as *ALOX5AP*, confers modest risk of myocardial infarction in an Icelandic cohort. Measurements of leukotriene B4 (LTB4) production suggest that this risk is mediated through upregulation of the leukotriene pathway. Three cohorts from the United States also show that HapK confers a modest relative risk (1.16) in European Americans, but it confers a threefold larger risk in African Americans. About 27% of the European American controls carried at least one copy of HapK, as compared with only 6% of African American controls. Our analyses indicate that HapK is very rare in Africa and that its occurrence in African Americans is due to European admixture. Interactions with other genetic or environmental risk factors that are more common in African Americans are likely to account for the greater relative risk conferred by HapK in this group.

To search for SNPs and potential causal variants of *LTA4H*, we sequenced DNA across the *LTA4H* gene region (42 kb) in 93 individuals affected with myocardial infarction. Although no coding sequence variant leading to amino acid substitutions was found, we identified and selected eight SNPs and genotyped them, together with two known SNPs in the 5′ region of the gene (Fig. 1), in Icelandic individuals with myocardial infarction and controls. These SNPs extend 11.9 kb upstream and 1 kb downstream of the *LTA4H* coding sequence and were selected to capture all haplotypes with a frequency of >2% across the gene region.

We tested the ten SNPs for association with myocardial infarction by using 1,553 individuals with myocardial infarction and 863 population-based controls. No single SNP or haplotype defined by the ten SNPs was found to be significantly more common in all individuals with myocardial infarction than in controls (Supplementary Tables 1 and 2 online). Therefore, we tested association of the haplotypes with more severe myocardial infarction phenotypes—namely, early-onset myocardial infarction and myocardial infarction with other cardiovascular diseases, including peripheral vascular disease, stroke, or both. Early-onset myocardial infarction did not show significant association with any of the haplotypes (data not shown); however, myocardial infarction with additional cardiovascular diseases showed association with a haplotype that we called HapK (Fig. 1 and Table 1). The frequency of HapK in individuals with myocardial infarction and additional cardiovascular disease and in controls was 14.5% and 10.4%, respectively, corresponding to a relative risk of 1.45 ($P = 0.0091$) for each copy of HapK carried ($P = 0.035$ after adjusting for the number of haplotypes tested).

To investigate the functional relevance of HapK, we examined the correlation between HapK carrier status and the amount of LTB4, the main product of the LTA4H enzyme, that was produced by granulocytes isolated from the same individuals. We have previously reported[1] that granulocytes from individuals with myocardial infarction ($n = 41$) produce more LTB4 than those from controls without any history of myocardial infarction ($n = 36$). This data set included 14 HapK carriers: seven individuals with myocardial infarction (one homozygote) and seven controls. Using multiple regression including age, gender and disease status as covariates, we observed a positive correlation between HapK and LTB4 production after stimulating the cells for 15 min ($P = 0.015$) and 30 min ($P = 0.009$) with ionomycin (Table 3 and Supplementary Table 3 online).

[1]deCODE Genetics, Inc., Sturlugata 8, IS-101Reykjavik, Iceland. [2]Laboratory of Genomic Diversity, National Cancer Institute, Frederick, Maryland 21702, USA. [3]National University Hospital, Reykjavik, Iceland. [4]Emory University School of Medicine, Atlanta, Georgia, USA. [5]University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [6]Icelandic Heart Association, Holtasmári 1, 201 Kópavogur, Reykjavik, Iceland. [7]Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA. Correspondence should be addressed to K.S. (kstefans@decode.is).

Received 8 June; accepted 5 October; published online 10 November 2005; doi:10.1038/ng1692

EXHIBIT H

LETTERS

Given the modest risk conferred by HapK in Iceland, we performed a replication study in three independent myocardial infarction cohorts from the United States recruited in Philadelphia, Cleveland and Atlanta. All three cohorts contained both self-reported European Americans and African Americans (Table 1), who were analyzed separately. Table 1 shows the association results for HapK in each of these cohorts. The *P* values reported for all of the replication analyses are one sided because we tested only HapK for increased risk. An excess of HapK was detected in European American individuals with myocardial infarction from Philadelphia (relative risk = 1.37, $P = 0.0051$) and Cleveland (relative risk = 1.12, not significant), but not in those from Atlanta (Table 1). The association of HapK with myocardial infarction in European Americans was replicated when the three cohorts were simply combined (relative risk = 1.19, $P = 0.006$), and when a Mantel-Haenszel–like[2] analysis was done to adjust for differences in HapK frequency among controls in the three cohorts (relative risk = 1.16, $P = 0.019$; Table 2). As in Iceland, the risk of HapK was greater in those individuals with myocardial infarction who had a history of stroke or peripheral vascular disease (Table 1), with the combined cohort adjusted analysis yielding a relative risk of 1.31 ($P = 0.037$; Table 2).

Although HapK was found to be less frequent in African Americans (Table 1), its association with myocardial infarction was much stronger in this group, with the relative risk estimated as 6.50, 1.78 and 5.21 for the cohorts from Philadelphia, Cleveland and Atlanta, respectively (Table 1). The estimated relative risk was substantially less in Cleveland than in the other two cohorts, mainly because the control frequency of HapK is greater in that cohort. The relative risk conferred by HapK in the combined group of all African Americans with cohort adjustment was estimated to be 3.57 ($P = 0.000022$). Its confidence interval did not overlap with that of the European Americans (Table 2), showing that the relative risk of HapK in these two groups is significantly different ($P < 0.001$).

As HapK is much more frequent in European Americans than in African Americans, it is possible that the greater relative risk of myocardial infarction in African Americans is in part attributable to a greater European ancestry in individuals with myocardial infarction than in controls. This could be caused either by a bias in data collection (such as differences in recruitment of the myocardial infarction and control groups), or because European ancestry itself is a risk factor for myocardial infarction in African Americans or a close correlate of such a risk factor. To investigate this further, we genotyped a set of 75 unlinked microsatellite markers, selected as informative for distinguishing between African and European ancestry (see Methods and Supplementary Table 4 online), in the three US cohorts, in 364 Icelanders and in 90 Nigerian Yorubans used in the HapMap project[3]. We used Structure software[4,5] to analyze these data to estimate the distribution of European ancestry in individuals grouped by disease status and self-reported ethnicity (Table 4). We also obtained estimates of European ancestry by applying a weighted least-squares (WLS) estimator[6] to a subset of the microsatellite alleles that showed the greatest differences in frequency between European and African populations in accordance with ref. 7 (Table 4). Overall, we found a close correspondence between self-reported ethnicity and the estimated ancestry derived from the genetic markers and also between the estimated individual ancestry (Structure) and group ancestry (WLS). In particular, the almost perfect assignment of African ancestry to Nigerian Yorubans and European ancestry to Icelanders indicated that the admixture estimates of the American individuals with myocardial infarction and controls were reliable. Furthermore, our estimates of European ancestry in African Americans were in the range reported in most previous studies[7–11].

Notably, we found that African American individuals with myocardial infarction had, on average, a slightly greater European ancestry than did the African American controls in the Philadelphia and Atlanta cohorts (Table 4). When all three cohorts were combined, the African American individuals with myocardial infarction and controls had on average 22.3% and 19.9% European ancestry, respectively (one-sided $P = 0.046$). This difference can largely be accounted for by a few individuals who were recorded as African Americans but had a relatively large European ancestry. We corrected for potentially misclassified individuals by excluding from the study self-reported African Americans with <20% African genetic ancestry according to the Structure results (seven individuals with myocardial infarction and four controls). The result was a notable reduction in the difference between individuals with myocardial infarction (20%) and controls (19.2%). Controlling for ancestry, whether by excluding potentially misclassified individuals or by using individual European ancestry estimates as covariate[12], referred to as 'admixture adjustment', has a negligible effect on the relative risk and statistical significance of the association of HapK with myocardial infarction in African Americans (Tables 1 and 2). We conclude that the higher relative risk of HapK in African Americans is not simply a consequence of differences in European ancestry between individuals with myocardial infarction and controls.

Notably, however, African American carriers of HapK had, on average, more European ancestry than those who did not carry HapK: 28.9% versus 19.8% (two-sided $P = 0.00008$). This is consistent with the observation that HapK was not found in the Nigerian HapMap sample, but was relatively common in the Icelandic and the CEPH CEU (Utah residents with ancestry from northern and western Europe) samples used in the HapMap project (Supplementary Fig. 1 and Table 2 online). Although HapK was found to be common in the Asian HapMap samples, the Structure-based estimate of Asian ancestry in African Americans was small (~1%), supporting the hypothesis that copies of HapK present in African Americans are mostly of European origin. Furthermore, we detected no difference in Asian ancestry between African American individuals with myocardial infarction and controls or between HapK carriers and noncarriers.

The *LTA4H* gene is located in a single linkage disequilibrium (LD) block in both European and African populations and is the only gene known in that block (Supplementary Fig. 2 online). To identify a single causal variant captured by HapK, we sequenced a region of 75 kb

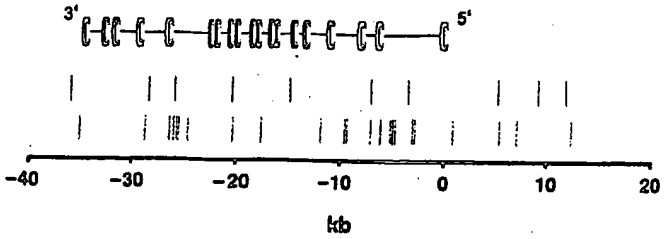


Figure 1 Structure of the *LTA4H* gene. Exons are shown as pink cylinders, and the positions of all genotyped SNPs relative to exons are shown as green lines. The SNPs and alleles (defined on the plus strand) defining HapK are SG12S16 (C), rs2660880 (G), rs6538697 (T), rs1978331 (A), rs17677715 (T), rs2247570 (T), rs2660898 (T), rs2540482 (C), rs2660845 (G) and rs2540475 (G). See information on SG12S16 in Supplementary Table 1. The relative position of SNPs typed in the HapMap project[3] (Phase I, version 16c.1) are shown as gray lines. For Icelanders and European Americans, the association results in Tables 1 and 2 could be obtained with only five SNPs (rs1978331, rs17677715, rs2540482, rs2660845 and rs2540475). For African Americans, because of admixture effects, two more SNPs (rs2247570, rs2660898) had to be added to the above five to reproduce the results obtained with HapK.

LETTERS

encompassing the LD block containing *LTA4H* in several pooled DNA samples of Icelandic individuals with myocardial infarction and controls. Some pooled samples contained only HapK carriers. In addition, we examined the correlation of HapK with other SNPs in the HapMap[3] database (Phase I, version 16c.1). The best single SNP surrogate of HapK identified through both of these approaches was rs2660899 ($R^2 = 0.7$ in the CEU samples). We genotyped this SNP in the Philadelphia cohort, in which HapK showed the strongest effect. Although the T allele conferred a relative risk of 1.31 ($P = 0.008$) in European Americans, it did not fully capture the disease association with HapK in this African American

Table 1 Association of HapK with myocardial infarction

	Frequency of HapK			
Cohorts (*n*)	Individuals with MI	Controls	Relative risk	*P* value[a]
Icelanders				
All MI (1,553/863)	0.113	0.104	1.1	0.36
MI and additional CVD (325/863)	0.145	0.104	1.45	0.0091
European Americans				
Philadelphia				
All MI sre (728/430)	0.186	0.143	1.37	0.0051
All MI gda (724/430)	0.186	0.143	1.37	0.0051
All MI admix adj			1.36	0.0048
Cleveland				
All MI sre (627/792)	0.166	0.151	1.12	0.15
All MI gda (626/792)	0.166	0.151	1.11	0.16
All MI admix adj			1.12	0.15
MI and additional CVD sre (144/792)	0.193	0.151	1.34	0.046
MI and additional CVD admix adj			1.34	0.044
Atlanta				
All MI sre (236/553)	0.135	0.143	0.94	0.64
All MI gda (236/553)	0.135	0.143	0.94	0.64
All MI admix adj			0.94	0.63
MI and additional CVD sre (39/553)	0.173	0.143	1.25	0.25
MI and additional CVD admix adj			1.24	0.26
African Americans				
Philadelphia				
All MI sre (105/127)	0.103	0.017	6.5	0.000067
All MI gda (100/126)	0.104	0.018	6.45	0.000088
All MI admix adj			6.34	0.00010
Cleveland				
All MI sre (53/111)	0.122	0.072	1.78	0.11
All MI gda (52/111)	0.112	0.072	1.61	0.17
All MI admix adj			1.75	0.11
MI and additional CVD sre (13/111)	0.152	0.072	2.31	0.14
MI and additional CVD admix adj			2.27	0.16
Atlanta				
All MI sre (39/149)	0.075	0.015	5.21	0.018
All MI gda (38/146)	0.071	0.016	4.71	0.025
All MI admix adj			5.08	0.019
MI and additional CVD sre (8/149)	0.202	0.015	16.36	0.0039
MI and additional CVD admix adj			16.67	0.0035

Shown is the frequency of HapK in individuals with myocardial infarction (MI) and controls, together with the corresponding numbers (*n*) of subjects (individuals with myocardial infarction/controls), the relative risk and *P* values. Results are shown for European Americans and African Americans, defined by their self-reported ethnicity (sre). For each self-reported group, results are also shown for those who had a genetically detected ancestry (gda) of at least 20% European (in European Americans) and at least 20% African (in African Americans). Results adjusted for admixed ancestry in each self-reported group are also shown (admix adj). Myocardial infarction and additional cardiovascular diseases (CVD) refer to those individuals with myocardial infarction who also had either peripheral vascular disease or who had suffered a stroke. Information on previous history of stroke or peripheral vascular disease was not available for the subjects from Philadelphia.

[a]*P* values are two-sided for Icelanders but one-sided in all the other cohorts because we specifically tested the excess of HapK in individuals with myocardial infarction relative to controls.

LETTERS

Table 2 Association of HapK with myocardial infarction in combined American cohorts

Ethnic groups (*n*)	Frequency of HapK		RR (95% CI)	*P* value	PAR
	Individuals with MI	Controls			
European Americans					
All MI (1,591/1,775)	0.171	0.148	1.19 (1.04, 1.36)	0.006	
All MI coh adj			1.16 (1.01, 1.34)	0.019	
All MI coh adj, admix adj			1.16 (1.01, 1.33)	0.017	0.046
MI and additional CVD (183/1345)[a]	0.192	0.15	1.35 (1.00, 1.81)	0.026	
MI and additional CVD coh adj			1.31 (0.97, 1.78)	0.037	
MI and additional CVD coh adj, admix adj			1.32 (0.98, 1.78)	0.035	0.089
African Americans					
All MI (197/387)	0.105	0.032	3.52 (1.96, 6.29)	0.000012	
All MI coh adj			3.57 (1.94, 6.57)	0.000022	
All MI coh adj, admix adj			3.50 (1.90, 6.43)	0.000029	0.144
MI and additional CVD (21/260)[a]	0.176	0.041	4.94 (1.58, 15.43)	0.003	
MI and additional CVD coh adj			4.39 (1.32, 14.64)	0.008	
MI and additional CVD coh adj, admix adj			4.17 (1.21, 14.30)	0.012	0.219

The results describe the association of HapK with myocardial infarction (MI) in combined groups of self-reported European and African Americans from Philadelphia, Cleveland and Atlanta. The haplotype frequencies, the relative risk (RR) and the *P* values are shown first without any population adjustment; second, after adjusting for different cohort or population frequencies (coh adj); and third, after further adjusting for the admixture of African and European ancestries in each ethnic group (admix adj). All *P* values are one sided. PAR is the population attributable risk. CI, confidence interval.

[a]Only the Cleveland and Atlanta cohorts were combined for the severe phenotype of myocardial infarction and additional cardiovascular disease, as this information was not available for the subjects from Philadelphia.

cohort (**Supplementary Fig. 3** online). Thus, rs2660899 can be ruled out as a sole causal variant captured by HapK.

In theory, the observed association of myocardial infarction with HapK could be the result of an association with a causal variant located in the neighborhood of *LTA4H* but outside the LD block. Such a situation might explain the high relative risk observed in the recently admixed African Americans, potentially boosted by strong admixture-derived LD, and the modest relative risk in the nonadmixed groups of European Americans and Icelanders. Given the existing patterns of LD in European and African populations, however, the kind of admixture found in African Americans, which we examined by creating a 4:1 mixture of haplotypes from the Yoruban and CEPH CEU HapMap samples, would not be expected to produce a correlation ($R^2 > 0.25$) between HapK and any known SNP outside the *LTA4H* LD block. Because the observed effect of HapK on myocardial infarction is very strong in African Americans, it is implausible that the association is the consequence of a variant that is only loosely correlated with HapK. In addition, in an analysis of five markers located just outside the *LTA4H* LD block with significant allele frequency differences between African and European American controls, none was associated with HapK or differed between African American individuals with myocardial infarction and controls (**Supplementary Table 5** online). Thus, the difference in ancestry between African American individuals with myocardial infarction and controls seems to be localized to HapK.

The identification of a genetic variant that confers such different risks of myocardial infarction in African Americans and populations of European descent suggests a strong interaction between HapK and other genetic variants and/or non-genetic risk factors that are more common in African Americans than in European Americans and Icelanders. Our results emphasize that although genetic differences between human continental groups are small[13,14], some of these differences may nonetheless contribute to ethnicity-based health disparities[15], whether through frequencies of risk alleles, through risk conferred by such alleles, or both. We and others[16] have found a strong correspondence between self-reported ethnicity and genetically estimated ancestry. However, ancestry is a quantifiable trait, particularly in heterogeneous or recently admixed populations such as African Americans, that needs to be assessed to interpret reliably interactions among ancestry, genes and environment in the pathogenesis of disease[11,17,18].

Several reports indicate that the leukotriene pathway has a role in the pathogenesis of atherosclerosis, in particular in the branch involved in LTB4 biosynthesis[19–21]. We have shown that HapK is correlated with risk of myocardial infarction and increased production of LTB4, the main product of the enzyme encoded by *LTA4H*. LTB4 produced through activation of the leukotriene pathway may amplify inflammatory responses in the arterial wall, by mediating chemotaxis and thereby promoting adhesion of leukocytes to the vascular endothelium and transmigration. In addition, LTB4-induced activation of leukocytes leads to the release of lysosomal enzymes such as myeloperoxidase and the generation of reactive oxygen species[22], which have been attributed to initiation, propagation and acute complications of atherosclerosis[23,24]. Overall, these findings suggest that agents affecting LTB4 biosynthetic pathways may prove useful for primary or secondary prevention of heart attacks.

METHODS

Subjects from Iceland. The study cohort comprised 1,553 unrelated Icelandic subjects with myocardial infarction, including 597 with early-onset myocardial infarction and 325 with additional atherosclerotic manifestations (stroke and/or peripheral arterial disease), and 863 unrelated population controls. Recruitment

LETTERS

Table 3 Correlation between LTB4 and myocardial infarction and HapK carrier status[a]

	After 15 min	After 30 min
Predictor variable[a]	*P* value	*P* value
Disease status[b]	0.011	0.016
Carriers of HapK	0.015	0.009

[a]Two-sided *P* values correspond to a correlation between LTB4 after ionomycin stimulation of isolated granulocytes and both myocardial infarction status and the carrier status of the at-risk haplotype HapK. The results for the two time points were calculated by multiple regression with age, sex, disease status and carriers status as predictor variables and log-transformed LTB4 quantities as the response.

[b]The correlation between disease status and LTB4 has been reported previously[1].

of the cohort has been described previously[1]. In brief, individuals with myocardial infarction were recruited from a registry that includes all individuals with myocardial infarction diagnosed before the age of 75 in Iceland from 1981 to 2002, according to WHO-MONICA criteria for acute myocardial infarction[25]. Neurologists and vascular surgeons confirmed the diagnoses of stroke and peripheral vascular disease, respectively.

The Data Protection Commission and the National Bioethics Committee of Iceland approved the study. Informed consent was obtained from all study participants. Personal identifiers were encrypted with a third-party encryption system[26].

Subjects from Philadelphia. Study participants were enrolled at the University of Pennsylvania Medical Center through the PENN CATH study program, which studies the association of biochemical and genetic factors with coronary artery disease in subjects undergoing cardiac catheterization. In total 3,850 subjects have participated. For our study, we selected from the PENN CATH study 833 individuals (728 European Americans and 105 African Americans) diagnosed with myocardial infarction on the basis of either criteria for acute myocardial infarction (an increase in cardiac enzymes and electrocardiographic changes) or a self-reported history of myocardial infarction. For controls, we selected 557 individuals (430 European Americans and 127 African Americans) who showed no evidence of coronary artery disease (luminal stenosis less than 10%) on coronary angiography. Ethnicity information was self-reported.

The University of Pennsylvania Institutional Review Board approved the study, and all subjects provided written informed consent.

Subjects from Cleveland. Study participants were enrolled at the Cleveland Clinic Heart Center through the Genebank program, which is a registry of data and biological samples obtained from individuals undergoing coronary catheterization. The diagnostic criteria for myocardial infarction were based on at least two of the following: prolonged chest pain, electrocardiogram patterns consistent with acute myocardial infarction or a significant increase in cardiac enzymes. Subjects from the Genebank registry who lacked both significant luminal stenosis (<50% stenosis), as assessed by coronary angiography, and a previous history of coronary artery disease were selected as controls for the current study.

The study group included 680 individuals with myocardial infarction (627 European Americans and 53 African Americans) and 903 controls (792 European Americans and 111 African Americans). Ethnicity information was self-reported.

The study was approved by the Cleveland Clinic Foundation Institutional Review Board on Human Subjects, and all subjects gave written informed consent.

Subjects from Atlanta. Study participants were enrolled at the Emory University Hospital, the Emory Clinic and Grady Memorial Hospitals through the Emory Genebank and Clinical Registry in Neurology (CRIN). The Emory Genebank studies the association of biochemical and genetic factors with coronary artery disease in subjects undergoing cardiac catheterization. So far, 736 subjects have participated. For our study, those subjects who had a self-reported history of myocardial infarction (236 European Americans and 39 African Americans) were selected for the myocardial infarction group. Control subjects (553 European Americans and 149 African Americans) were selected from a group of individuals with nonvascular neurological diseases (mainly Parkinson and Alzheimer diseases) recruited from CRIN, their spouses, unrelated friends and community volunteers. These subjects were matched for age and ethnicity to the population with myocardial infarction population. Controls were excluded if they had a known history of myocardial infarction. All subjects provided written informed consent. Information on ethnicity was self-reported.

Statistical analysis. The haplotype association study was done with the program NEMO[27], which handles missing genotypes and uncertainty with phase through a likelihood procedure using the expectation-maximization algorithm to estimate haplotype frequencies. We emphasize that the likelihood ratio tests used explicitly take the uncertainty of the haplotypes counts into consideration, distinguishing them from a two-step procedure that first estimates haplotype counts and then treats the estimated counts as though they are actual counts. The relative risk of a particular haplotype was calculated by a multiplicative model in which the risk of the two alleles of a haplotype that a person carries multiplies[28,29]. With cohort adjustment, the model used for testing was essentially the Mantel-Haenszel test[2], in which each cohort is allowed to have different control haplotype frequencies, but the relative risk is assumed to be the same across cohorts. We extended the standard Mantel-Haenszel test to take into account the incomplete information on haplotype counts. Our admixture adjustment was similar to that proposed in ref. 12, in which the baseline or control frequencies of haplotypes are assumed to be a function of the admixture fraction and a likelihood ratio test is used. Similar to the Mantel-Haenszel model, however, we assumed here that the relative risk is a constant independent of admixture fraction, whereas it is assumed otherwise in ref. 12. The difference is likely to be small here, as we did the admixture adjustment separately in self-reported African Americans and in self-reported European Americans, and not in a combined group.

We used the program Structure[5] to estimate the genetic ancestry of individuals. Structure infers the allele frequencies of K ancestral populations on the basis of multilocus genotypes from a set of individuals and a user-specified value of K, and it assigns a proportion of ancestry from each of the inferred K populations to each individual. Our data set was analyzed by the admixture model, in which the ancestry prior alpha was allowed to vary among populations. This is an important option when genetic material from the K inferred ancestral populations (in this case the African and European ancestral populations) is not equally represented in the data set. This was clearly the case in our data set, which contained 3,366 self-reported European Americans, 584 self-reported African Americans, 364 Icelanders and 87 Nigerians. We ran Structure several times for each value of K in the range 2 to 5. We used the Icelanders and European Americans to identify the European ancestry component in the African Americans and the Nigerians to identify the African ancestry component. On the basis of these runs, we found evidence to indicate that $K = 3$ provides the best estimates of European ancestry in African Americans.

First, these estimates corresponded closely to independent group estimates based on Long's WLS admixture estimator[6]. Second, the results obtained with $K = 3$ indicated the existence of clearly defined African and European ancestral gene pools and a third gene pool that contributed a small amount (1–2%) to European and African Americans but nothing to Nigerians and Icelanders. An independent Structure analysis that also included Native American and East Asian reference samples indicated that this third component represented Asian ancestry. When $K > 3$, the European component became divided equally among the additional ancestral gene pools, whereas the African and Asian components remained stable in single components. Thus, $K > 3$ did not seem to provide any additional resolution to the data. Notably, the estimates of European ancestry for African American individuals were strongly correlated between different runs of Structure, regardless of the value of K. Thus, the average Spearman's rank correlation between runs was 0.987 and had a minimum of 0.964. The statistical significance of the difference in mean European ancestry between African American individuals with myocardial infarction and controls was evaluated by reference to a null distribution derived from 10,000 randomized data sets.

To genetically evaluate ancestry of the study cohorts from the US, we selected 75 unlinked microsatellite markers (**Supplementary Table 4** online) from about 2,000 microsatellites genotyped in a multiethnic cohort of 35 European

LETTERS

Table 4 Distribution of genetically determined European ancestry in myocardial infarction case-control cohorts

Cohort	Self-reported ethnicity	Disease status	WLS group estimate of European ancestry (s.e.m.)[a]	Distribution of estimated individual European ancestry[b]			
				Mean	s.d.	Median	25th–75th percentile range
Yoruban Nigerians	African	N/A	N/A	0.036	0.024	0.03	0.019–0.043
Iceland	European	N/A	N/A	0.991	0.015	0.994	0.990–0.996
All American	Eur. Am.	Individuals with MI	0.98 (0.0083)	0.965	0.083	0.991	0.977–0.995
All American	Eur. Am.	Controls	0.979 (0.0079)	0.969	0.07	0.992	0.979–0.995
Philadelphia	Eur. Am.	Individuals with MI	0.974 (0.0101)	0.955	0.101	0.99	0.971–0.995
Philadelphia	Eur. Am.	Controls	0.969 (0.009)	0.959	0.09	0.991	0.969–0.995
Cleveland	Eur. Am.	Individuals with MI	0.982 (0.0079)	0.971	0.068	0.991	0.980–0.995
Cleveland	Eur. Am.	Controls	0.981 (0.0081)	0.972	0.06	0.991	0.979–0.995
Atlanta	Eur. Am.	Individuals with MI	0.995 (0.0075)	0.981	0.038	0.991	0.984–0.994
Atlanta	Eur. Am.	Controls	0.982 (0.0092)	0.973	0.066	0.993	0.983–0.995
All American	Afr. Am.	Individuals with MI	0.243 (0.0138)	0.223	0.184	0.178	0.108–0.282
All American	Afr. Am.	Controls	0.213 (0.016)	0.199	0.145	0.174	0.094–0.267
Philadelphia	Afr. Am.	Individuals with MI	0.252 (0.0178)	0.235	0.195	0.188	0.121–0.288
Philadelphia	Afr. Am.	Controls	0.213 (0.0217)	0.186	0.137	0.157	0.082–0.257
Cleveland	Afr. Am.	Individuals with MI	0.232 (0.0222)	0.21	0.174	0.16	0.096–0.282
Cleveland	Afr. Am.	Controls	0.239 (0.0219)	0.223	0.136	0.191	0.127–0.281
Atlanta	Afr. Am.	Individuals with MI	0.226 (0.0246)	0.206	0.166	0.167	0.098–0.283
Atlanta	Afr. Am.	Controls	0.198 (0.0128)	0.193	0.155	0.161	0.086–0.252

[a]Long's WLS measure of admixture[6] was calculated with alleles from the set of 75 microsatellite markers. Frequencies from Icelanders and Nigerians were used to represent the ancestral allele frequencies of the European and African parental gene pools, respectively. In line with ref. 7, only 16 loci with alleles showing large differences in frequency ($\delta \geq 0.5$) between the two parental populations were used. For the African American cohorts, we calculated the WLS admixture statistic using the European American controls from the same city as representatives of the ancestral European gene pool. In each case, the estimate of European ancestry was higher by about 0.01 (data not shown). This is likely to be due to the small fraction of African alleles present in the European Americans and indicates that Icelanders serve as effective representatives of the European component of the European American gene pool. The WLS admixture statistic was also calculated by using alleles of all 75 microsatellite markers, yielding estimates of European ancestry in African Americans that were slightly higher than those reported above (by 0.01–0.02).

[b]Estimates of genetic ancestry were obtained from the Structure software using the parameters and data described in Methods. Note that the output from Structure does not label the ancestral admixture proportions as either 'European', 'African' or 'Asian', but rather as 'inferred cluster 1', 'inferred cluster 2' or 'inferred cluster 3'; however, the distribution of ancestry from these inferred clusters in Icelanders, Nigerians and the American cohorts suggests that they have a relatively straightforward correspondence with the labels 'African', 'European' and 'Asian' ancestry.

Americans, 88 African Americans, 34 Chinese and 29 Mexican Americans[30]. Out of the 2,000 microsatellite markers, the selected set showed the most significant differences among the European Americans, African Americans and Asians, and also had good quality and yield. Thirty-one of these markers have been used for similar purposes elsewhere[16].

Accession codes. GenBank: *LTA4H*, NM_000895.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the participants who made this study possible; the nurses at the Icelandic Heart Association; personnel at the deCODE core facilities; A. Rosen, T. Kamineni, J. Pareira and the CRIN staff for recruiting subjects; J. Pritchard for discussion on the admixture analyses; and members of the International HapMap Consortium for providing data which were crucial for our analysis. This work was supported by grants from the US National Institutes of Health (NIH) and by the Emory Alzheimer's Disease Center. Research in Atlanta was supported by NIH grant U54 ES012068 and by Emory General Clinic Research Center grant MO1-RR00039. Research in Cleveland was supported by NIH grant P50 HL077107-01.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Helgadottir, A. *et al.* The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat. Genet.* **36**, 233–239 (2004).
2. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22**, 719–748 (1959).
3. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
4. Falush, D., Stephens, M. & Pritchard, J.K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
5. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
6. Long, J.C. The genetic structure of admixed populations. *Genetics* **127**, 417–428 (1991).
7. Parra, E.J. *et al.* Estimating African American admixture proportions by use of population-specific alleles. *Am. J. Hum. Genet.* **63**, 1839–1851 (1998).
8. Adams, J. & Ward, R.H. Admixture studies and the detection of selection. *Science* **180**, 1137–1143 (1973).
9. Chakraborty, R., Kamboh, M.I., Nwankwo, M. & Ferrell, R.E. Caucasian genes in American blacks: new data. *Am. J. Hum. Genet.* **50**, 145–155 (1992).
10. McKeigue, P.M., Carpenter, J.R., Parra, E.J. & Shriver, M.D. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann. Hum. Genet.* **64**, 171–186 (2000).
11. Reiner, A.P. *et al.* Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *Am. J. Hum. Genet.* **76**, 463–477 (2005).
12. Pritchard, J.K., Stephens, M., Rosenberg, N.A. & Donnelly, P. Association mapping in structured populations. *Am. J. Hum. Genet.* **67**, 170–181 (2000).
13. Bamshad, M., Wooding, S., Salisbury, B.A. & Stephens, J.C. Deconstructing the relationship between genetics and race. *Nat. Rev. Genet.* **5**, 598–609 (2004).
14. Jorde, L.B. & Wooding, S.P. Genetic variation, classification and 'race'. *Nat. Genet.* **36** (Suppl.), S28–S33 (2004).
15. Royal, C.D. & Dunston, G.M. Changing the paradigm from 'race' to human genome variation. *Nat. Genet.* **36**, S5–S7 (2004).
16. Tang, H. *et al.* Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *Am. J. Hum. Genet.* **76**, 268–275 (2005).
17. Barnholtz-Sloan, J.S., Chakraborty, R., Sellers, T.A. & Schwartz, A.G. Examining population stratification via individual ancestry estimates versus self-reported race. *Cancer*

LETTERS

Epidemiol. Biomarkers Prev. **14**, 1545–1551 (2005).
18. Hoggart, C.J. *et al.* Control of confounding of genetic associations in stratified populations. *Am. J. Hum. Genet.* **72**, 1492–1504 (2003).
19. Spanbroek, R. *et al.* Expanding expression of the 5-lipoxygenase pathway within the arterial wall during human atherogenesis. *Proc. Natl. Acad. Sci. USA* **100**, 1238–1243 (2003).
20. Aiello, R.J. *et al.* Leukotriene B4 receptor antagonism reduces monocytic foam cells in mice. *Arterioscler. Thromb. Vasc. Biol.* **22**, 443–449 (2002).
21. Mehrabian, M. *et al.* Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice. *Circ. Res.* **91**, 120–126 (2002).
22. Samuelsson, B. Leukotrienes: mediators of immediate hypersensitivity reactions and inflammation. *Science* **220**, 568–575 (1983).
23. Nicholls, S.J. & Hazen, S.L. Myeloperoxidase and cardiovascular disease. *Arterioscler. Thromb. Vasc. Biol.* **25**, 1102–1111 (2005).
24. Brennan, M.L. *et al.* Prognostic value of myeloperoxidase in patients with chest pain. *N. Engl. J. Med.* **349**, 1595–1604 (2003).
25. Joint International Society and Federation of Cardiology and World Health Organization Task Force on Standardization of Clinical Nomenclature. Nomenclature and criteria for diagnosis of ischemic heart disease. *Circulation* **59**, 607–609 (1979).
26. Gulcher, J.R., Kristjansson, K., Gudbjartsson, H. & Stefansson, K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur. J. Hum. Genet.* **8**, 739–742 (2000).
27. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat. Genet.* **35**, 131–138 (2003).
28. Falk, C.T. & Rubinstein, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann. Hum. Genet.* **51**, 227–233 (1987).
29. Terwilliger, J.D. & Ott, J. A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* **42**, 337–346 (1992).
30. Smith, M.W. *et al.* A high-density admixture map for disease gene discovery in African Americans. *Am. J. Hum. Genet.* **74**, 1001–1013 (2004).

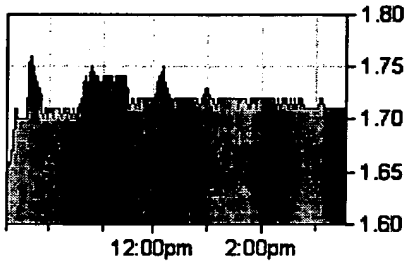
Monogram Provides Update on Trofile(TM) Co-Receptor Tropism Assay

Pfizer Launches Multi-National Expanded Access Program for Maraviroc

Dec 1, 2006 6:00:00 AM

SOUTH SAN FRANCISCO, Calif., Dec. 1 /PRNewswire-FirstCall/ -- Monogram Biosciences, Inc. (Nasdaq: MGRM) announced today that its collaborator Pfizer, Inc. (NYSE: PFE) has separately announced plans to establish a multi-national Expanded Access Program (EAP) that will make its investigational CCR5 antagonist maraviroc available to HIV/AIDS patients who have limited treatment options due to resistance or intolerance. Monogram's co-receptor tropism assay, Trofile, was used for patient selection for maraviroc's clinical development program, and the two companies are engaged in a collaboration agreement to make Monogram's assay available for patient use on a global basis.

Related Quotes



	Last	Change	% Change
MGRM	1.71 ▼	0.03	1.79
PFE	26.94 ▼	0.17	0.64

"This investigational therapy represents a potential milestone in the treatment of HIV," said Monogram CEO William Young. "We applaud Pfizer's vision to sculpt a new model for drug development that so closely integrates advanced diagnostics into the clinical program. We are proud to be a part of this ground-breaking effort."

Pfizer also confirmed plans to submit applications for marketing approval of maraviroc in both the U.S. and EU following review of the data from the two currently ongoing Phase 3 clinical trials of the drug. The company expects to submit these study results for presentation at an upcoming HIV conference.

Maraviroc is designed to work differently from other available HIV medications. CCR5 antagonists block the virus from gaining access into healthy cells via the CCR5 co-receptor, a common pathway for viral entry. Monogram's Trofile co-receptor tropism assay identifies whether individual strains of HIV use the CCR5 co-receptor, the CXCR4 co-receptor or both co- receptors to infect healthy cells. This helps clinicians determine whether a CCR5 antagonist like maraviroc may be a good therapeutic option for treating individual patients.

Pfizer's EAP is intended to provide access to maraviroc for patients who, in the opinion of the program investigators, have an urgent need for novel medicines because of viral resistance or intolerance to currently

Related News

Jan 12/07 07:00
Monogram Biosciences Pric
Placement of Convertible S
Unsecured Notes (PRN)

Jan 02/07 14:24
Monogram Biosciences To
the 25th Annual JPMorgan
Conference (PRN)

Dec 01/06 09:18
Hot Stocks to Watch for Fric
December 01, 2006: Expan
the $2.7 Trillion Global Tele
Industry! (MRW)

Dec 01/06 06:00
Monogram Provides Update
(TM) Co-Receptor Tropism
(PRN)

Nov 21/06 06:00
Monogram Biosciences to F
Personalized Medicine: A W
Opportunities (PRN)

Business News

Jan 17/07 15:32
Silver Quest Resources Ltd
Silver Discovery-Corcoran (
NV (CCN)

Jan 17/07 15:30
College of Teachers Will Se
Innovative Ways to Assess
Teaching Qualifications (CN

Jan 17/07 15:29
Start A Heart, Save A Life: S
and Heart and Stroke Foun
kickstart the campaign for d
in Ontario communities (CN

Jan 17/07 15:27
Skylon Funds Announce Di
for Month Ending January 3
(CNW)

Jan 17/07 15:24
Coalition of Canadian Ener
supports Finance Committe
to review trust tax legislatio

Jan 17/07 15:22
Meridex Announces the Ap
of New Board Member (CCI

Jan 17/07 15:20
President Leaving Buhler In
Inc. (CNW)

Jan 17/07 15:16
Blue Jays sign Fasano (CN

EXHIBIT 2

available therapies. To be eligible for the program, patients must be clinically stable with documented CCR5-tropic HIV-1 infection.

In a study presented at the International AIDS Conference in Toronto in August 2006 by scientists from Pfizer, the negative predictive value of Monogram's Trofile co-receptor tropism assay was assessed for maraviroc (Study 1029). Results show that patients identified by the assay as having virus using both the CXCR4 and CCR5 receptors (dual/mixed tropic) did not respond virologically to the investigational (CCR5) therapy. These data suggest that screening patients with the Trofile assay will allow physicians to optimize treatment regimens for their HIV patients.

Maraviroc and other entry inhibitors currently in development come at a time when increasing drug resistance makes treating HIV more difficult than ever. Highly sensitive and precise diagnostic tools are playing an ever more important role in the development of new therapeutic approaches that give new hope to physicians and patients running low on options.

About Monogram Biosciences, Inc.

Monogram is advancing individualized medicine by discovering, developing and marketing innovative products to guide and improve treatment of serious infectious diseases and cancer. The Company's products are designed to help doctors optimize treatment regimens for their patients that lead to better outcomes and reduced costs. The Company's technology is also being used by numerous biopharmaceutical companies to develop new and improved antiviral therapeutics and vaccines as well as targeted cancer therapeutics. More information about the Company and its technology can be found on its web site at http://www.monogrambio.com.

Forward Looking Statements

Certain statements in this press release are forward-looking. These forward-looking statements include references to the potential for an HIV drug that requires a molecular diagnostic for patient selection, the ability of the Company to advance its opportunities in HIV, and activities expected to occur in connection with the Pfizer collaboration. These forward-looking statements are subject to risks and uncertainties and other factors, which may cause actual results to differ materially from the anticipated results or other expectations expressed in such forward-looking statements. These risks and uncertainties include, but are not limited to: the risk that regulatory authorities may not require a molecular diagnostic for patient selection for an HIV drug, risks related to the implementation of the collaboration with Pfizer; risks related to progress Pfizer's clinical trial and any ultimate approval of maraviroc, risks related to our ability to recognize revenue from activities under the collaboration with Pfizer; risks and uncertainties relating to the performance of our products; the growth in revenues; the size, timing and success or failure of any clinical trials for CCR5 antagonists, entry inhibitors or integrase inhibitors; the use of our Trofile co-receptor tropism assay for patient use in the event of approval of any CCR5 antagonists; our ability to successfully conduct clinical studies and the results obtained from those studies; whether larger confirmatory clinical studies will confirm the results of initial studies; our ability to establish reliable, high-volume operations at commercially reasonable costs; expected reliance on a few customers for the majority of our revenues; the annual renewal of certain customer agreements; actual market acceptance of our products and adoption of our technological approach and products by pharmaceutical and biotechnology companies; our estimate of the size of our markets; our estimates of the levels of demand for our products; the impact of competition; the timing and ultimate size of pharmaceutical company clinical trials; whether payors will authorize reimbursement for our products and services; whether the FDA or any other agency will decide to further regulate our products or services; whether we will encounter problems or delays in automating our processes; the ultimate validity and enforceability of our patent applications and patents; the possible infringement of the intellectual property of others; whether licenses to third party technology will be available; whether we are able to build brand loyalty and expand revenues. For a discussion of other factors that may cause our actual events to differ from those projected, please refer to our most recent annual report on Form 10-K and quarterly reports on Form 10-Q, as well as other subsequent filings with the Securities and Exchange Commission. We do not undertake, and specifically disclaim any obligation, to revise any forward-looking statements to reflect the occurrence of anticipated or unanticipated events or circumstances after the date of such statements.

Trofile is a trademark of Monogram Biosciences, Inc.

contacts: Alfred G. Merriweather
Chief Financial Officer
Tel: 650 624-4576
amerriweather@
monogrambio.com

Jeremiah Hall
Feinstein Kean Healthcare
Tel: 415 677-2700
jeremiah.hall@
fkhealth.com

SOURCE Monogram Biosciences, Inc.

Jan 17/07 15:16
The École supérieure de ba
contemporain on the road fo
recruitment - 2007 National
Tour (CNW)

Jan 17/07 15:11
Rolling Rock Resources An
10,000 Metre Drill Program
Monument Bay (CCN)

Alfred G. Merriweather
Chief Financial Officer of Monogram Biosciences
Inc.
+1-650-624-4576
or amerriweather@monogrambio.com; or Jeremiah Hall of Feinstein Kean Healthcare
+1-415-677-2700
or jeremiah.hall@fkhealth.com

San Francisco Business Times - August 21, 2006
http://sanfrancisco.bizjournals.com/sanfrancisco/stories/2006/08/21/story13.html

SAN FRANCISCO Business Times

Studies bolster Monogram Biosciences' AIDS therapy test

San Francisco Business Times - August 18, 2006 by Daniel S. Levine

A South San Francisco diagnostics company is getting a boost from studies presented at an international AIDS conference that show the importance of its tests for new therapies.

The studies, scheduled to be presented at the XVI International AIDS Conference in Toronto on Aug. 17, show that the Monogram Biosciences' assay can determine whether a patient would benefit from a new class of AIDS drugs by detecting which co-receptor the virus uses to enter and hijack a cell.

HIV can invade cells either through the CCR5 co-receptor or the CCRX4 co-receptor or both. Pfizer is in a late-stage clinical trial for Maraviroc, a drug that blocks entry to the CCR5 receptor and is on target to apply for approval to market the drug with U.S. regulators before year end and has been using Monogram's test in conjunction with trials of its drug.

"Drug resistance continues to be a major problem in HIV management, and patients are in need of new classes of drugs, including CCR5 antagonists," said Monogram CEO Bill Young. "Despite advances in treatment options for HIV-infected patients, we know that not every drug candidate is appropriate for every patient. Our assays help screen patients to identify those most likely to respond to these new classes of drugs based on the tropism of the infecting virus."

Monogram's test can show whether the virus in a particular patient is using CCR5, CCRX4 or both to determine whether a patient would benefit from the drug. The test, which will likely cost between $1,000 and $1,500, would also be needed to monitor a patient using the drug because even though the virus uses CCR5 in 80 percent of early HIV cases, the virus can mutate and use CCRX4 or both co-receptors.

Nate Cornell, biotechnology analyst for Pacific Growth Equities, said the Monogram test could reach peak annual sales of as much as $30 million a year with Maraviroc's approval. As other similar drugs come to market, he said demand for Monogram's test would increase.

"It gives investors confidence in the Monogram assay that if Pfizer's drug Maraviroc is approved, Monogram's test will be a required tool to start the therapy," said Cornell of the studies. "It's one of the most exciting examples of the role of personalized medicine on the market today."

In May, Monogram and Pfizer reached a non-exclusive alliance to make Monogram's test available globally. The companies did not release details of the agreement, but it involved a $25 million investment by Pfizer in Monogram and makes Pfizer responsible for sales of the assay outside of the United States.

Monogram will provide the assay to Pfizer at a fixed price and process the tests at its lab in South San Francisco. Pfizer will also cover any costs associated with setting up the systems and support of the distribution, processing and support of the assays outside of the United States.

Daniel S. Levine covers biotechnology for the San Francisco Business Times.